DR112839

TECH. MEMO
P 1179

COPY

AD-A220 351

ROYAL AEROSPACE ESTABLISHMENT

HANDBOOK OF UNCERTAINTY METHODOLOGY FOR ENGINE TESTING AT PYESTOCK

by

J. C. Ascough

November 1989

DTIC
ELECTE
APR 11 1990
S D
E

Procurement Executive, Ministry of Defence
Farnborough, Hampshire

UNLIMITED

90 04 09 245

CONDITIONS OF RELEASE

0062742 BR-112839

MR PAUL A ROBEY
DTIC
Attn:DTIC-FDAC
Cameron Station-Bldg 5
Alexandria
VA 22304 6145
USA

********** DRIC U

COPYRIGHT (c)
1988
CONTROLLER
HMSO LONDON

********** DRIC Y

Reports quoted are not necessarily available to members of the public or to commercial organisations.

ROYAL AEROSPACE ESTABLISHMENT

Technical Memorandum P 1179

Received for printing 8 November 1989

HANDBOOK OF UNCERTAINTY METHODOLOGY FOR ENGINE TESTING AT PYESTOCK

by

J. C. Ascough

SUMMARY

Measurements of aircraft gas turbine engine performance in the altitude test facility at RAE(P) are subject to a small amount of uncertainty resulting from a combination of precision (or random) errors and bias (or systematic) errors. The limits of the precision errors can be readily calculated by statistical analysis of the results measured during the engine tests. Bias limits are not directly observable in the test results, but can be predicted by a comprehensive assessment of all possible sources of error, which are propagated to the test results.

Many people find these methods difficult to comprehend and apply and this Memorandum has been written for their benefit. It is a guide not only for engine test staff at Pyestock but also for their customers who need to be assured of the rigorous attention given to identifying and reducing measurement uncertainty.

Copyright
©
Controller HMSO London
1989

A-1

LIST OF CONTENTS

1 INTRODUCTION

Engine performance measurements in the altitude test facility at Pyestock are expected to be of 'good accuracy'. This is a qualitative feeling engendered by the knowledge that most of the instrumentation is calibrated and the performance results are scrutinised to catch any bad points as they come off the computer during the engine tests.

From time to time a complete quantitative assessment of engine test uncertainty (a better technical word than 'accuracy') is made. This may be a contractual requirement for a new engine project, or it might be needed for the calibration of a flight engine for which the calculated uncertainty would be propagated to the aircraft performance in flight. Another instance might be for 'trouble-shooting' to investigate the cause of an anomalous performance result so as to identify the most important sources of error where corrective action may be applied for best effect.

This Memorandum sets out in fairly simple terms the main features of the Uncertainty Methodology used at Pyestock. It is largely based on the methodology of Abernethy and Thompson[1], with some simplification and change of emphasis to suit local practices and resources. The methodology has also been extended in one respect to cover graphical effects by which a performance curve is shifted due to bias errors propagated through the horizontal axis.

Some of the material of the present report comes from experience with the Uniform Engine Test Programme (UETP)[2-4], in which RAE(P) participated with other international facilities to test and compare results from a pair of J57 engines.

Some of the phraseology may be unfamiliar, or have special meanings, but it is important that this be clearly understood. The reader is offered a preliminary exposure to some of the main terms in the Glossary, immediately ahead of the Notation. A fuller explanation is given in the main text. Subsequently, it is hoped, the Glossary will serve as a convenient aide memoire for future use.

2 UNCERTAINTY MODELS

Fig 1 indicates the notional model by which random (precision) and bias (systematic) errors occur. In the mind's eye there is a true value whose position or value is unknown. A fixed bias error, β, lies somewhere within the range of $\pm B$, the bias limits which are centred on the unknown true value. The bias interval is $2B$.

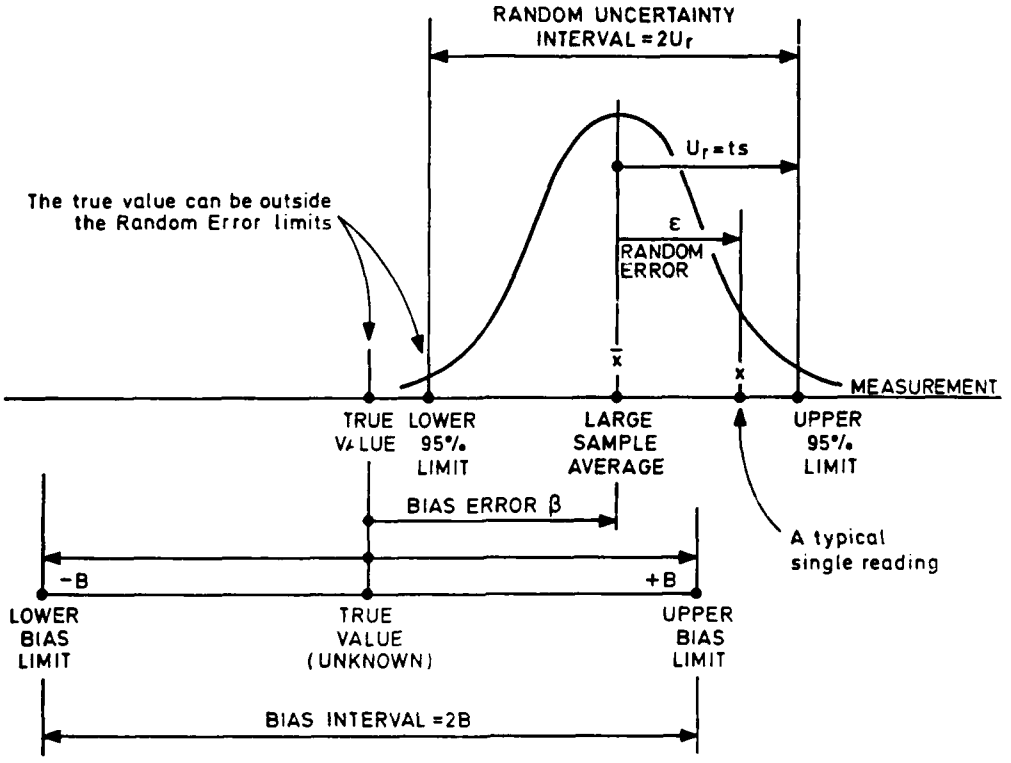


Fig 1 Notional model of random and bias errors

Random errors, such as ε, are scattered around the bias error, following a 'Normal' or 'Gaussian' distribution. The experimental standard deviation, s, of this distribution can be calculated from a sample of n points:

$$s = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} . \qquad (1)$$

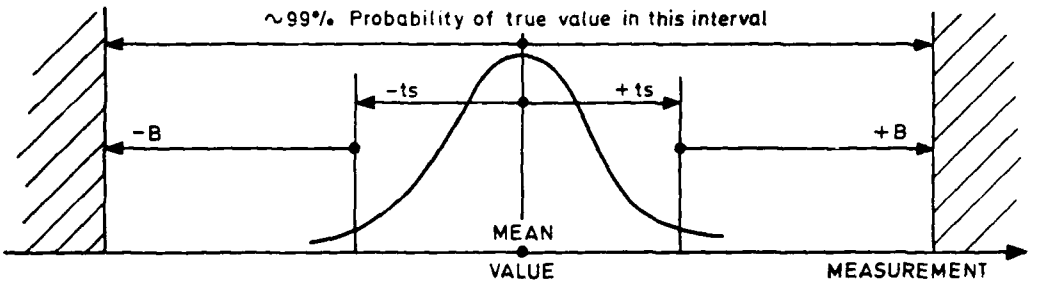
Thence the random uncertainty U_r (also known as the 95% confidence limits), is given as:

$$\pm U_r = \pm t_{95} s , \qquad (2)$$

where t_{95} is Student's "t", given by statistical tables. It is expected that 95% of the data points will fall within the limits $\pm U_r$.

The notation model is only used to establish the concept. Since the true value is always unknown, it is necessary to work with two practical models, described below.

ADDITIVE MODEL $U_{ADD} = (B + ts)$



ROOT SUM SQUARES MODEL $U_{RSS} = \sqrt{B^2 + (ts)^2}$

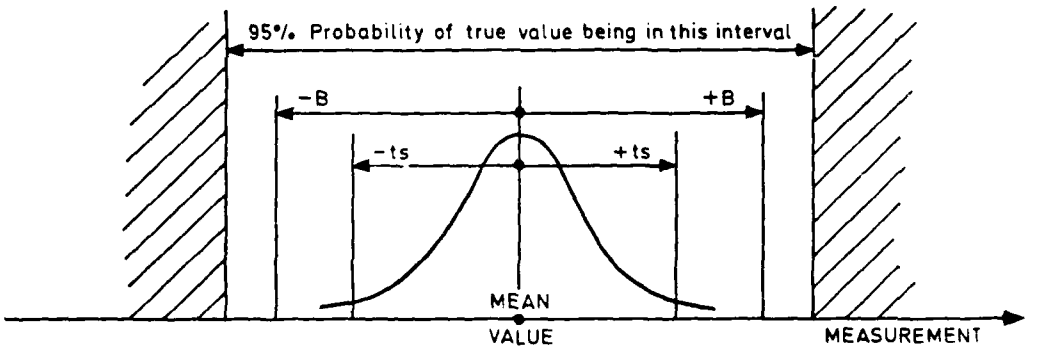


Fig 2 Practical uncertainty models

Fig 2 illustrates two practical uncertainty models which can be used. Both of these are centred on the observed mean value, which is always known, unlike the true value. With the 'additive model', viz:

$$\pm U_{add} = \pm (B + ts) , \qquad (3)$$

the bias limits $\pm B$ are simply added to the outer ends of the random uncertainty limits $\pm ts$.

The 'root-sum-squares' model is:

$$\pm U_{rss} = \pm \sqrt{B^2 + (ts)^2} . \qquad (4)$$

This model is more difficult to visualise, but it implies a probability distribution for bias errors as well as for random errors which are combined as independent components. For further discussion see Ref 5.

Abernethy has shown by Monte Carlo computer simulations, reported in Appendix D of Ref 1, and confirmed on the Pyestock computer, that $\pm U_{add}$ usually produces 99% coverage of experimental points, whereas $\pm U_{rss}$ produces 95% coverage.

There has traditionally been a preference for U_{rss} in the UK and Europe, but a preference for U_{add} in the USA. Either, or both, models can be used because an important feature of the prediction synthesis methodology is that B and (ts) components must be kept and reported separately.

The above discussion on bias and precision errors relates to a general measurement, x, illustrated conventionally in Fig 1 as a value on the horizontal axis. Similar considerations apply to graphical values of y such as SFC plotted against a correlating result, z, such as FN as illustrated in Fig 3. Here, the precision errors are taken to be the values in the y-direction scattered about the curve fit. The residual standard deviation, RSD takes the place of the precision index, s. The curve fit (CF) value of y supersedes the mean value, $\bar{x}$, and the separation of CF from the true value (TV) is the bias error, β. As in Fig 1, the 'notional distribution' of bias errors is centred on the true value, with the curve fit expected to lie somewhere between the limits (TV + B) and (TV - B). But, since the true value is unknown, the 'practical distribution' of bias errors is centred on the curve fit, with the true value expected to lie somewhere within the limits (CF + B) and (CF - B).

The statistical properties of a mean value and of a curve fit are explained in section 5.

3 THE DEFINED TEST RESULT AND MEASUREMENT PROCESS

The behaviour of errors in real life is very complicated, and so it is necessary to set up simplified models before any calculations can be made. In the past, many different models have been used, which has created great confusion. An ingenious way out of the difficulty is provided by an important feature of the Abernethy methodology[1], known as the 'Defined Measurement Process' (DMP). This was adopted by RAE(P) in the early 1980s to enable the previous 3-class system of MIDAP[6] to be converted into the simpler 2-class system of Abernethy[1].

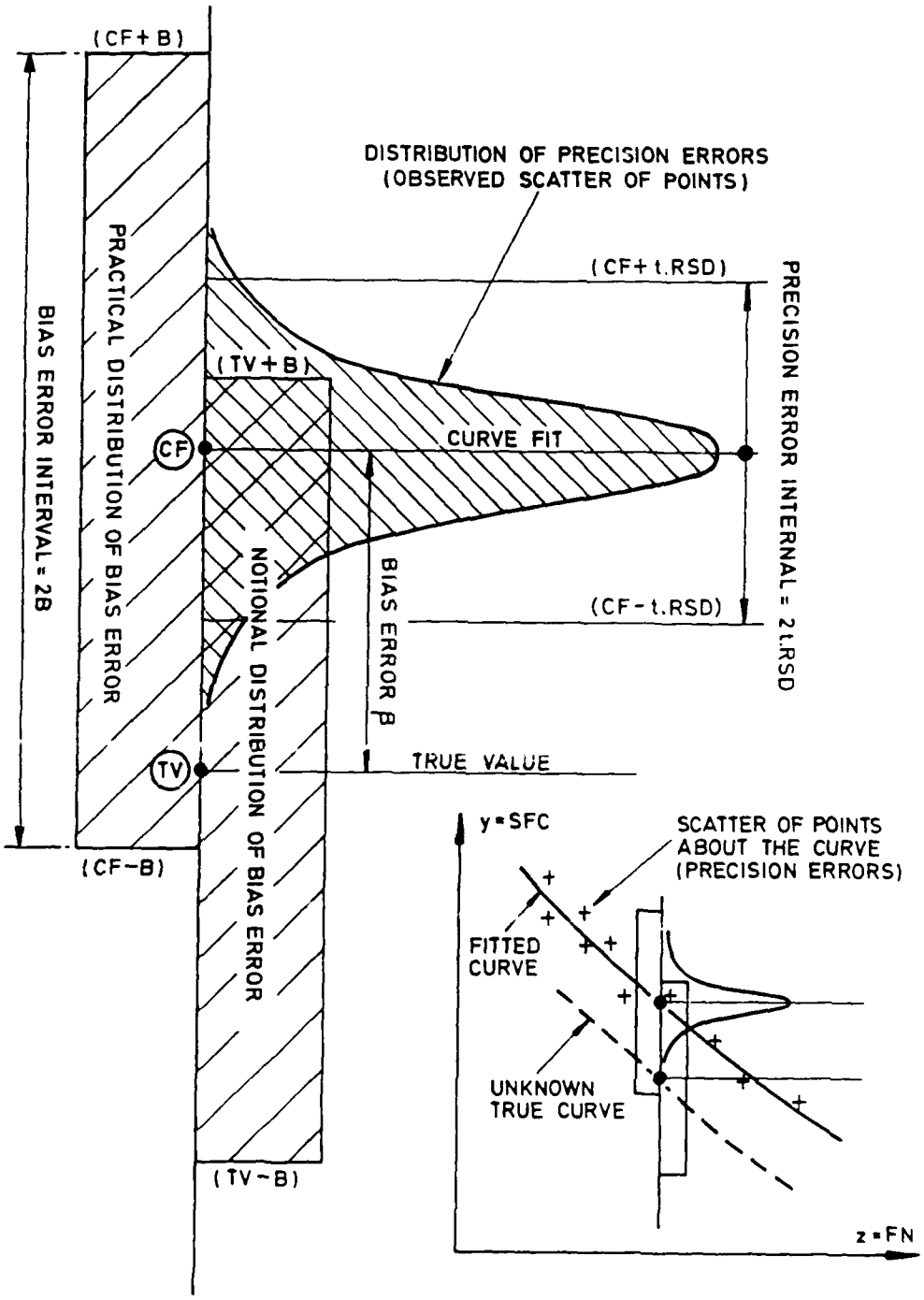


Fig 3 Bias and precision errors of a single SFC curve

To apply the DMP, it is first necessary to declare exactly what is the 'Defined Test Result' of the measurements. Then any error which appears as visible scatter in the defined test result is classified as 'Precision', and any error which remains constant in the defined result is classified as 'Bias'. The usual practice at RAE(P) is to declare the defined result to be a single performance curve, such as SFC vs FN, obtained in a single run in the test cell. Hence the scatter about the curve is regarded as indicative of precision errors, while the hidden constant errors which displace the curve away from the true curve are regarded as bias errors.

The way in which errors, from every possible source, propagate through the whole of the measurement system, is briefly described in the following sections. It is often found that the errors which behave as precision errors at some stage of the measurement process, end up as fixed errors in the defined result - this is known as 'error fossilisation' because the previously-live errors have died out to leave fixed, or dead, bias errors in the result. An example of this is the random scatter which blurs the position of an instrument calibration curve. Once such a calibration curve is accepted, the difference between it and the true curve becomes a fixed error, ie bias.

In the UETP exercise[2-4] some of the other participating facilities adopted different DMPs from that used by RAE(P), and so their classifications of precision and bias were different from those of RAE(P). For example, AEDC's precision class embraced errors which showed random variation at any time throughout the whole of the UETP exercise in their facility, including the links with their National Standards. One advantage of the RAE(P) DMP is that precision errors, in the form of observed scatter about the performance curves, are completely amenable to standard statistical treatment in the form of regression analysis, which can be done on-line while the engine is running. This accounts for all the precision errors that actually exist without the possibility of any of them being overlooked. Another advantage is that the theory of the 'curve shift effect', explained in section 7, strictly only applies to the RAE(P) definition of bias.

4 DIRECTION OF APPROACH

The techniques of 'Actual Results Analysis' and 'Prediction Synthesis' are alternative directions of approach, as indicated in Fig 4.

'Actual Results Analysis' is an application of statistical methods as can be found in standard text books. Ref 7 is a popular introduction, Ref 8 is more rigorous but still readable, while Ref 9 contains details of the regression

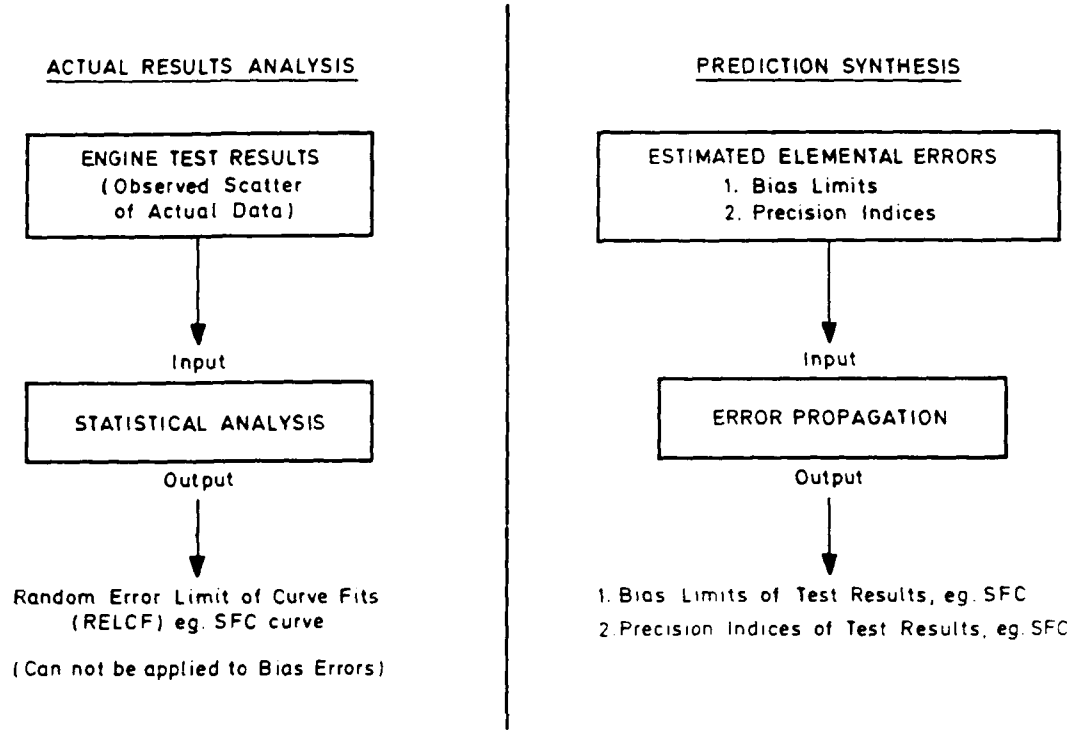


Fig 4 Alternative directions of approach for uncertainty assessment

analysis used to establish the random uncertainty of performance curve fits. Actual Results Analysis takes the evidence of data which actually occurs in the engine test to establish the Random Error Limit of Curve Fit (RELCF). This is that band within which the true position of a performance curve would lie if there were no bias errors. A fuller description is given in section 5.

The alternative direction of approach is known as 'Prediction Synthesis', which deals mainly with bias errors, but precision errors can also be included. The methodology of Abernethy and Thompson[1] is applied, with some slight modifications and changes of emphasis, to make uncertainty predictions. It is a process of synthesis because the complete answer is built up (ie synthesised) from estimates of the basic elements of uncertainty. A fuller description is given in section 6.

5 STATISTICAL ANALYSIS OF ACTUAL RESULTS

5.1 Properties of mean values and curve fits

A fundamental property of statistics (see Refs 7 to 9) is that the position of a mean value, $\bar{x}$, of 'n' Normally distributed points is established more tightly than a single point, x. Its experimental standard deviation is:

$$s(\bar{x}) = \frac{1}{\sqrt{n}} s(x) \quad . \tag{5}$$

This is indicated in the following sketch

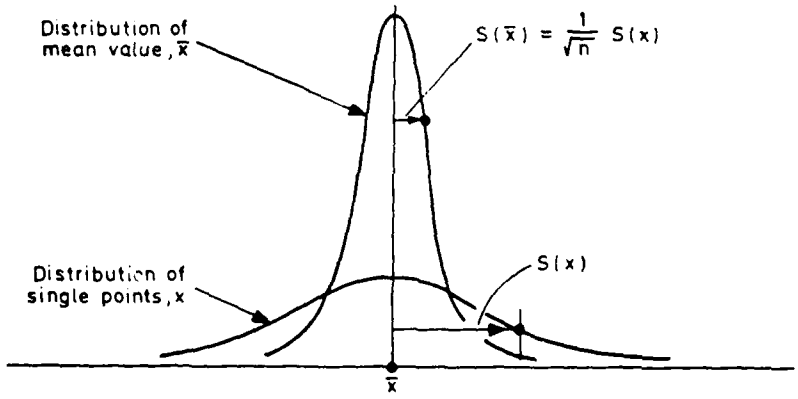


Fig 5 Probability distribution of a mean value

Equation (5) applies strictly to Normal (ie Gaussian) distributions. Fortunately, engine test data are usually close approximations to this - the Normality can be checked by statistical tests[7-9] if there is any doubt.

Engine test performance results are usually presented graphically in the form of one result, y (eg SFC) plotted against a correlating result, z (eg FN), and a curve $\hat{y}$ is fitted by the method of least squares (see Refs 7 to 9). So we need to deal with a curve fit, $\hat{y}$ instead of a mean value, $\bar{x}$ as indicated in the following Fig 6. Note that values along the horizontal axis (abscissa) as denoted by the symbol 'z' rather than 'x' because x_i is used for the input parameters, the errors of which are propagated to both axes, y and z.

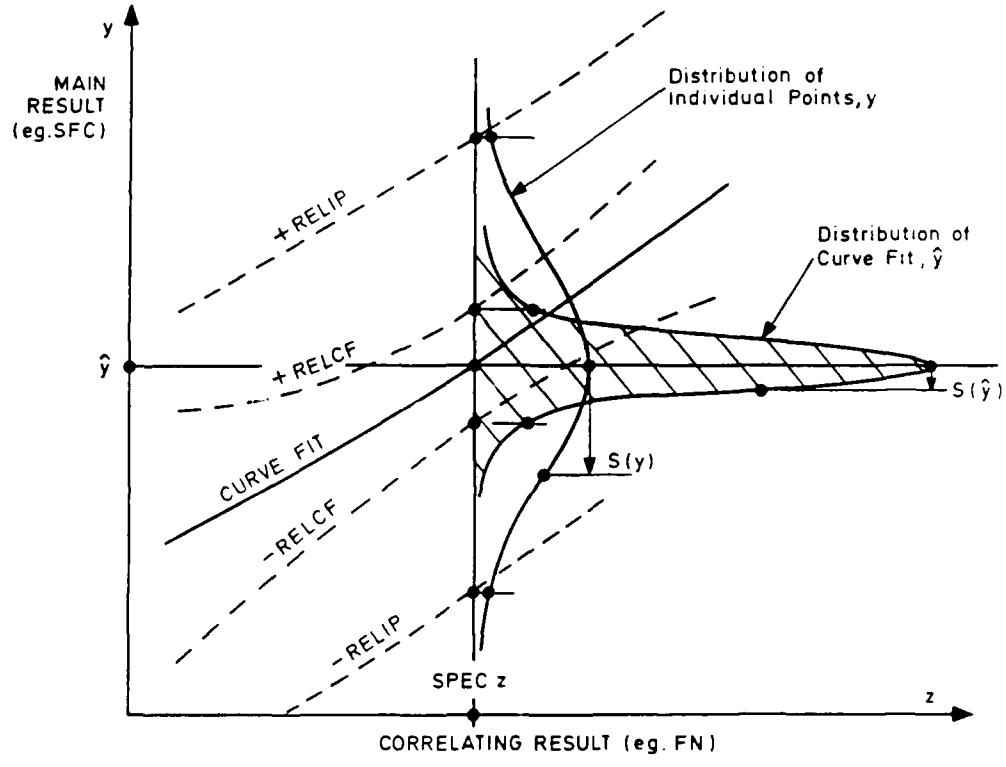


Fig 6 Probability distribution of a curve fit

At any specified value of z (SPEC Z) there is a curve fit value, $\hat{y}$. Its experimental standard deviation is:

$$s(\hat{y}) = RSD \left[\frac{1}{\sqrt{n}} + \text{other terms} \right]^{\frac{1}{2}}, \qquad (6)$$

where RSD is the residual standard deviation of the points scattered about the curve.

The complete expression for 'other terms' is given in Ref 9. Thus the position of $\hat{y}$ is established more tightly than that of a single point, y, and the improvement is related to the factor $1/\sqrt{n}$, rather like a mean value. The

95% confidence limits of $\hat{y}$, in other words the Random Error Limits of Curve Fit (RELCF) are given by:

$$\pm \text{RELCF} = \pm t_{95} s(\hat{y}) \quad . \tag{7}$$

These are shown by the inner dashed lines in Fig 6. They are quite close to the fitted curve in the centre of the graph, but they fan out at the ends of the data, becoming wide apart for the extrapolations. This latter effect is a useful safety feature.

Fig 7 shows a typical SFC curve fit from the UETP tests in Cell 3 at Pyestock. In this example the best fit was a quadratic curve, but sometimes a cubic curve fit is justified. At the SPEC Z value of 19.75 kN for FN, the curve fit value is $\hat{y}$ = 29.527 g kN^{-1} s^{-1} with the RELCF values ± 0.25%.

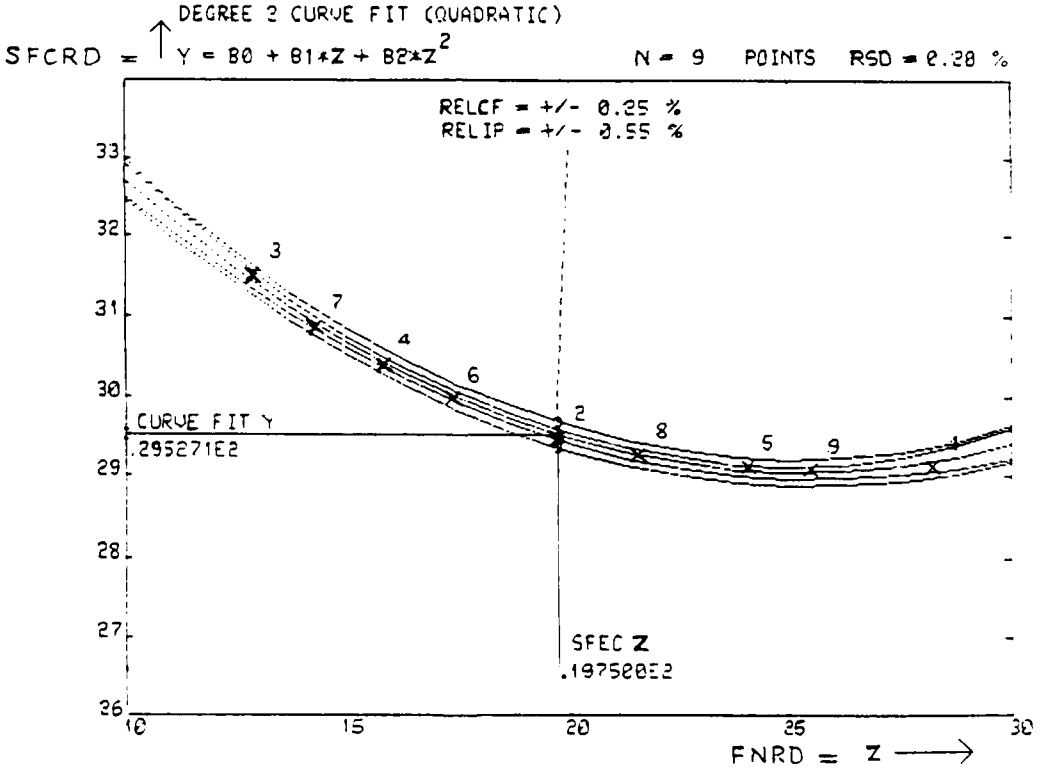


Fig 7 Typical SFC curve fit with RELCF and RELIP lines

For special 'Contract Tests', which must be agreed between all parties, these procedures can be incorporated in an Engine Test Code[10]. In this case the above analysis is developed so as to control the on-line testing in order to guarantee that an agreed random uncertainty is actually achieved. Moreover, this is done with the minimum testing to achieve that goal, by continuous computer analysis during the testing, taking just enough points (n) to ensure the required RELCF.

5.2 Random scatter and outlier detection

The statistical properties of mean values and curve fits described in the preceding section 5.1 assume that the surrounding scatter comes from Gaussian distributions. The diagrams of Figs 5 and 6 are very much idealised, as though these random errors were distributed smoothly. However, the real life behaviour of random errors is much more 'lumpy' than this. Even if there exists a well-defined background Gaussian distribution, it will be found that random samples taken from it are always somewhat irregular, especially with samples of small size. Appendix A shows the sort of experimental distributions that might arise with such random samples, from which mean values and curve fits are calculated.

In addition to the ragged nature of the actual values that might genuinely come from Gaussian distributions, there are sometimes a few 'outliers' superimposed that do not properly belong to the bulk of the data. It is important that these outliers be detected and eliminated from the mean values and curve fits. Procedures for doing this are fully described in Appendix B.

It is usual for the Grubbs tests to be incorporated in the computer data acquisition routines to automatically detect and eliminate outliers from the great number of mean values that are taken for a typical test point at Pyestock.

It is not the current practice to apply computer routines to detect outliers in the scatter about the engine performance curve fits, but a possible procedure is explained in section B.2.2 of Appendix B, whereby this might be done in the future.

The current practice is, briefly, as follows. If the position of the curve were known exactly, then the 95% confidence limits of the individual points about it would be given by:

$$\pm CL_{95}(y) = \pm t_{95} RSD \quad . \qquad (8)$$

A more rigorous expression, which takes into account the random uncertainty of the curve, is given by the Random Error Limit of Individual Points:

$$\pm \text{RELIP} = \pm [(t_{95}\text{RSD})^2 + \text{RELCF}^2]^{\frac{1}{2}} \quad (9)$$

and these lines are shown on Figs 6 and 7. Since RELCF decreases progressively as the number of points n increases, it follows that

$$\text{RELIP} \rightarrow t_{95}\text{RSD} . \quad (10)$$

Even with a modest number of points (say $n = 9$ as for the curve fit in Fig 7) it will be found that the value of RELIP is quite close to t_{95}RSD . With the actual data shown in Fig 7, the RELCF is 0.25%, the RSD is 0.20% and t_{95} is 2.45. Hence from equation (9)

$$\pm \text{RELIP} = \pm [(2.45 \times 0.20)^2 + 0.25^2]^{\frac{1}{2}} = \pm 0.55\%$$

and from equation (10)

$$\pm \text{RELIP} \rightarrow \pm 2.45 \times 0.20 = 0.49\%$$

which is quite a close approximation.

To complete the Pyestock procedure for outlier detection, we would expect and accept 1 point in 20 to fall just outside the ±RELIP boundaries, but any point much further away from the curve would be regarded as an outlier and hence a candidate for deletion. If deleted, the curve fit would then be recalculated with one less data point.

If this analysis is done on-line during the engine test, the opportunity exists for extra points to be taken to compare with the suspect point, to help the decision whether to reject or not. This procedure follows the principles of the Test Code[10] mentioned at the end of section 5.1 above.

6 PREDICTION SYNTHESIS

6.1 Preliminary sensitivity survey

It is usual to start with a preliminary 'sensitivity survey' to get some idea of the sensitivity of the various performance results (R_k) such as SFC, to the various 'basic measurements' (x_i) such as air meter static pressure (PSA). An alternative name for x_i is 'input parameter', because it is an input to the

engine performance calculations. This requires the calculation of the 'influence coefficients' θ_{ki} (sometimes called 'sensitivity factors'). In dimensional terms:

$$\theta_{ki} = \frac{\partial R_k}{\partial x_i} , \tag{12}$$

where ∂R_k is the change in the result R_k due to a change ∂x_i in an input. In principle, it should be possible to evaluate θ_{ki} by differential calculus, but the relationship between R_k and x_i in engine performance calculations is usually too complicated for this. Instead, θ_{ki} values can conveniently be found by perturbation of the computer calculations, providing that the engine performance program has been written. The calculation of these θ_{ki} values is itself done by computer program. It is more convenient to express the influence coefficients in relative form:

$$\theta'_{ki} = \frac{x_i}{R_k} \frac{\partial R_k}{\partial x_i} . \tag{13}$$

In practice, not only are different influence coefficients required for each x_i, for each R_k, but also for each flight condition. Thus

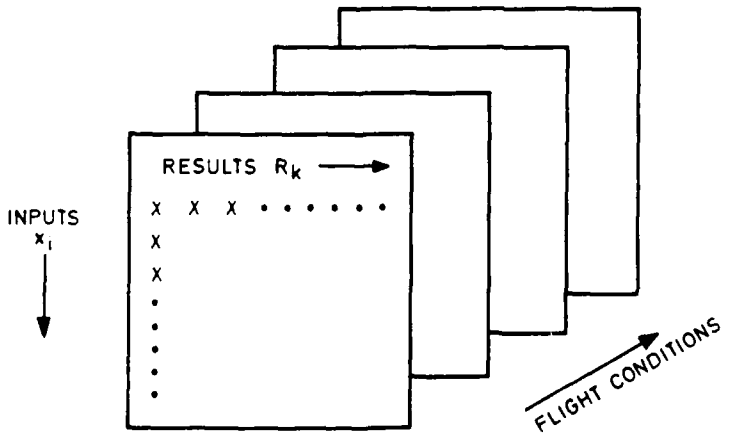


Fig 8 Matrix of influence coefficients

The central column of Fig 9 depicts typical influence coefficients.

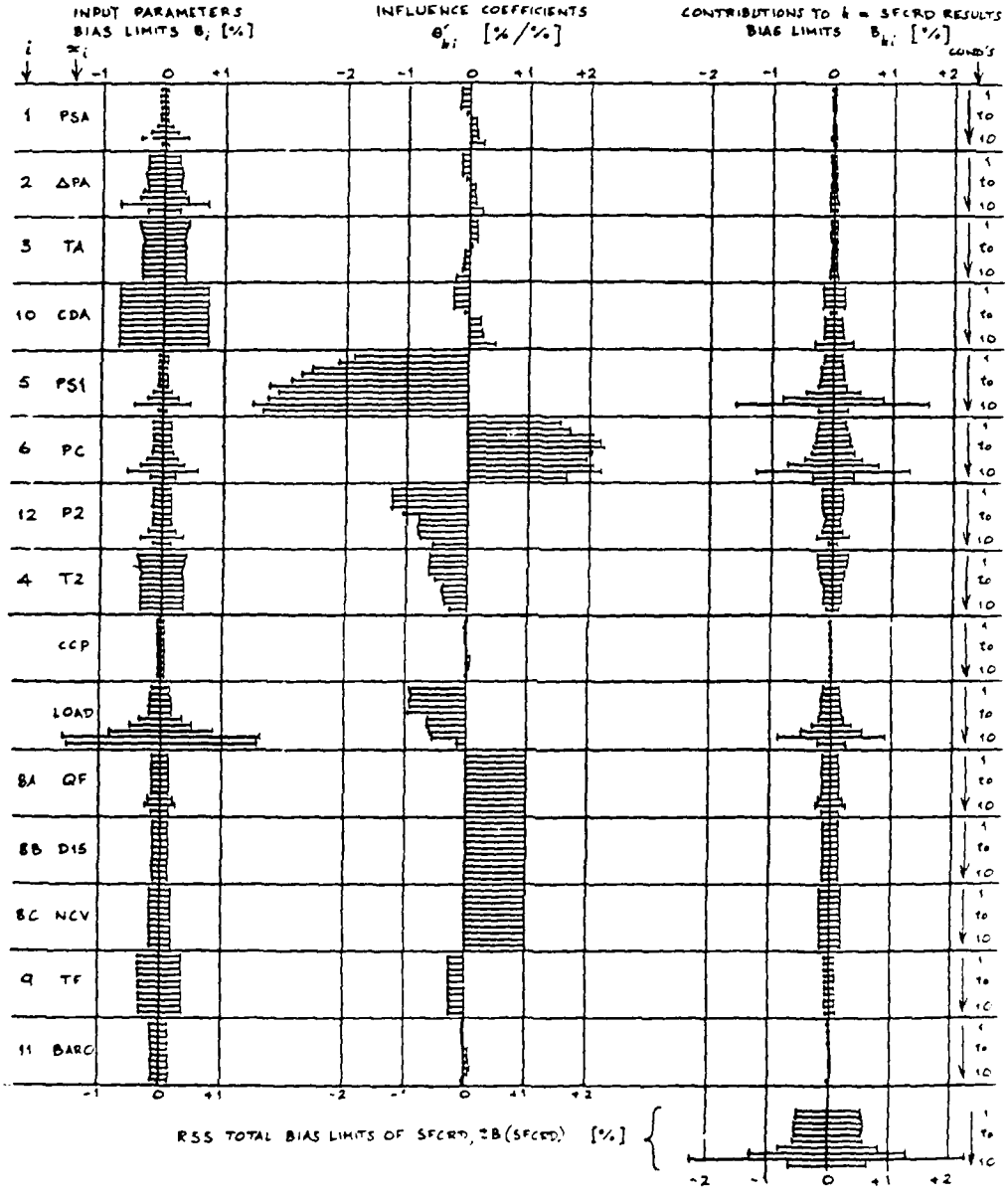


Fig 9 Propagation of bias limits from input parameters to (SFCRD) at target points (prediction synthesis for J57 UETP tests in Cell 3, RAE, Pyestock)

In this example, taken from the UETP exercise in Cell 3, the matrix consisted of 15 values of x_i , 4 values of R_k and 10 flight conditions - a total of 600 different influence coefficients. Perhaps the biggest problem here is to display this information in such a way that the reader can taken it in! The central column of Fig 9 depicts 150 of such influence coefficients in bar chart form, for the single result of R_k = SFCRD .

Having established these influence coefficients, one can see at a glance which are the most important input parameters where most attention must be given to keep the errors down. Fig 9 shows that x_5 = PS1 (the engine inlet static pressure) is the most important item for R_k = SFCRD . At Flight Condition 9 the influence coefficient is

$$\theta'_{k5} = -3.5\%/\% \tag{14}$$

that is, 1% error in PS1 will create -3.5% error in SFCRD. At the other extreme, it can be seen from Fig 9 that Barometer error has negligible effect on this particular result, with the worst influence coefficient at Flight Condition 8 of

$$\theta'_{k11} = 0.1\%/\% \tag{15}$$

that is, 1% error in Barometer will create only 0.1% error in SFCRD. The reason for this relative immunity to error in Barometer is the beneficial effect of error-cancellation in the complicated engine performance calculations for SFCRD.

(The left-hand and right-hand columns of Fig 9 are described in the following section.)

6.2 Complete prediction synthesis

The strategy is laid out in Fig 10. The basic measurements x_i (which were encountered in section 6.1 as input parameters for the Sensitivity Survey) are now the interface between Step 1 and Step 2.

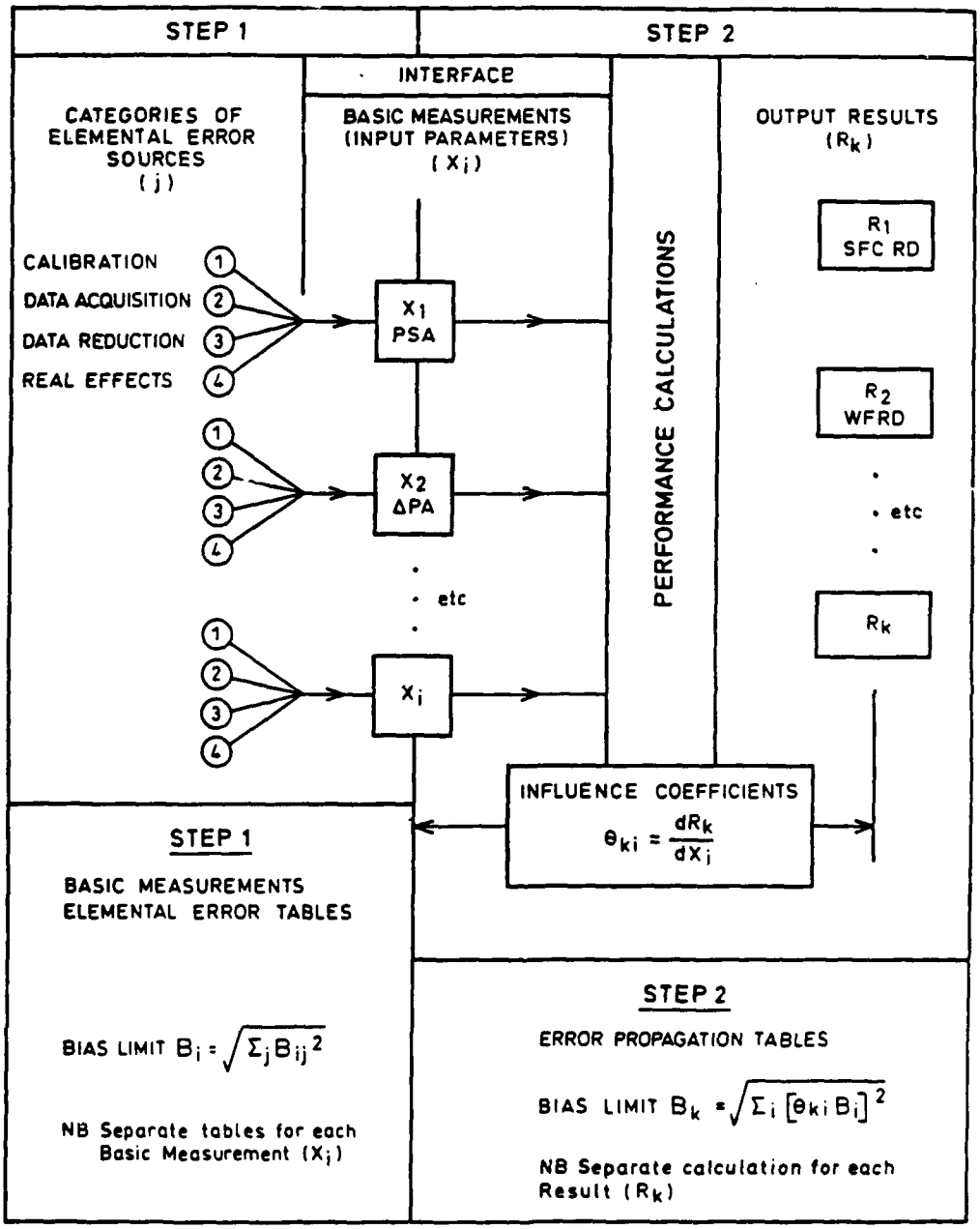


Fig 10 Error propagation through basic measurements interface

For Step 1, separate 'elemental error' tables are prepared for each x_i. Table 1 is an example, where x_7 is the test frame load. Error sources are assessed under four categories (j) as indicated in Fig 11.

	INSTRUMENTATION		NON-INSTRUMENTATION	
	ERROR CATEGORIES			
	1 CALIBRATION HIERARCHY	2 DATA ACQUISITION	3 DATA REDUCTION	4 REAL EFFECTS
ERROR SOURCES	1 NAT. STD 2 TFR. STD 3 LAB. STD 4 WKG. STD MEAS. INST.	1. EXCIT. VOLTAGE 2. ELEC. SIMUL'N 3. SIGNAL COND'N 4. TRANSDUCER 5. RECORDING (ALL OBVIATED BY END TO END CALIBRATIONS)	1. CURVE FITS 2. RESOLUTION (USUALLY NEGLIGIBLE)	1. SPATIAL AVERAGING 2. FLUCTUATIONS 3 TEST FRAME MECHANICS

Fig 11 Elemental error sources (for each basic measurement)

Category 1 covers the calibration hierarchy, traceable all the way from the measuring instrument (a Bofors shear force load cell in Table 1) to the National Standard at NPL. At each stage the calibration consists of comparing one standard with another. Fig 12 shows how the original biases are removed by these calibrations, but in their place there remains a combination of random uncertainty with a total standard deviation for the complete calibration hierarchy of

$$s_{CAL} = \sqrt{s_1^2 + s_2^2 + s_3^2 + s_4^2} \,. \qquad (16)$$

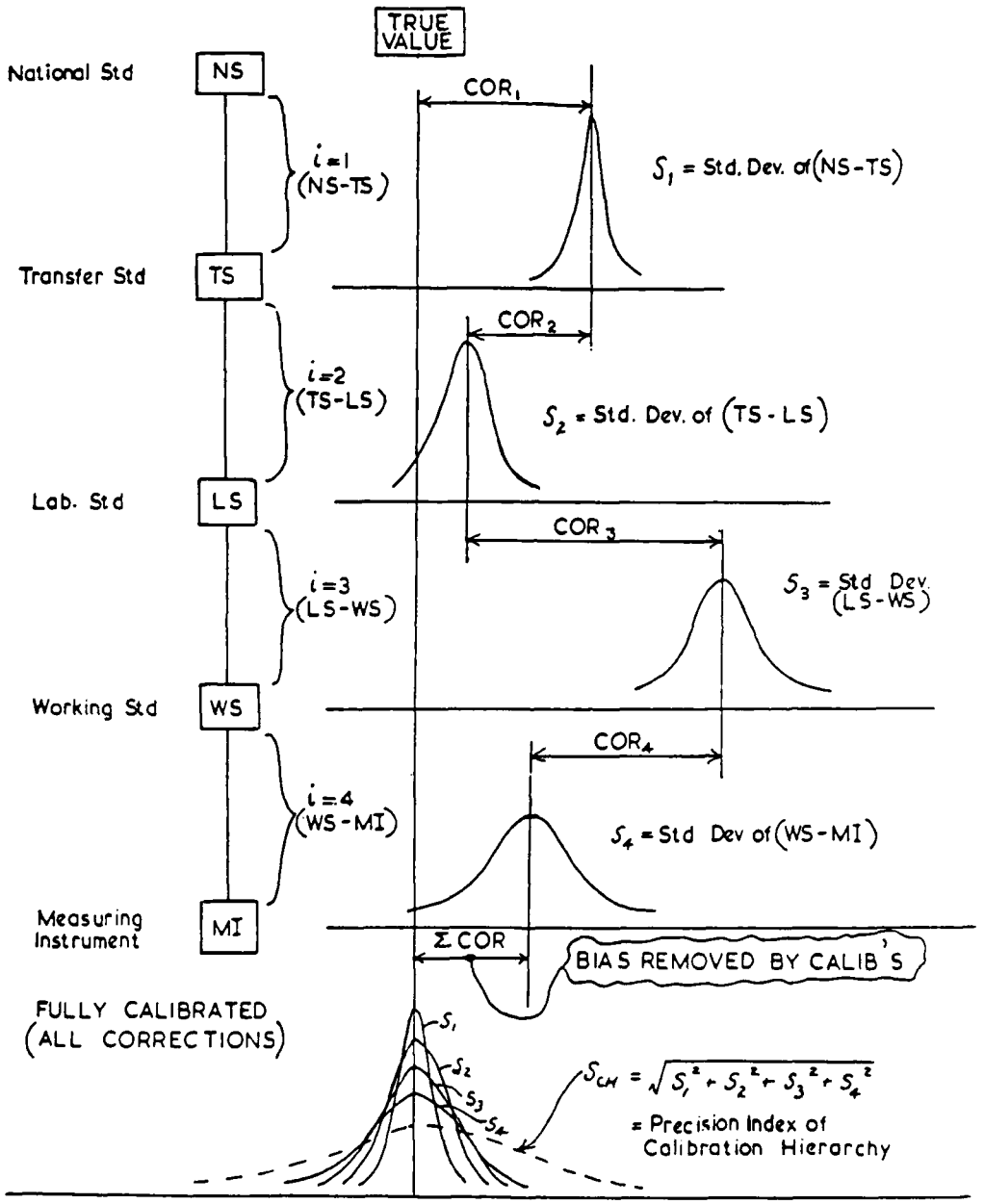


Fig 12 Original biases removed by calibration (replaced by calibration uncertainty)

However, the residual errors from this calibration hierarchy are propagated to the engine test as a single fixed error, not as random scatter, ie they are 'fossilised' from random errors into bias errors within the limits

$$B_1 = ts_1 , \quad B_2 = ts_2 , \quad \text{etc} . \tag{17}$$

Hence

$$B_{CAL} = \sqrt{B_1^2 + B_2^2 + B_3^2 + B_4^2} . \tag{18}$$

Fig 13 illustrates how any one stage of the calibration hierarchy may be analysed. The standard deviation in this case would be

$$s_{\text{one stage}} = \sqrt{\sum_{i=1}^{5} \frac{(\hat{y}_i - \bar{\hat{y}})^2}{5 - 1}} , \tag{19}$$

where $\hat{y}_i$ is the ith curve fit.

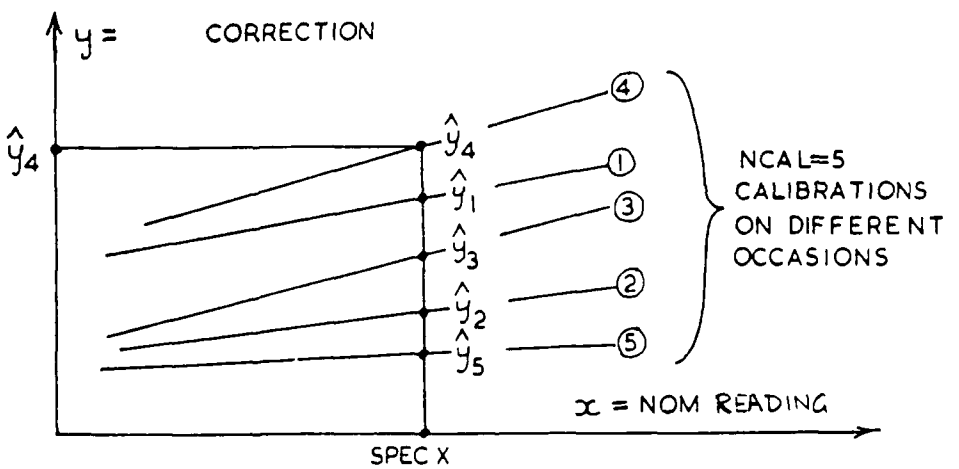


Fig 13 Statistical analysis of different calibration lines

In the case of test frame load in Table 1, the calibration hierarchy is quite short, with only two stages plus a small basic uncertainty of the National

Standard. These elemental bias limits are entered separately in Table 1, because they will be root-sum-squared down the complete table.

Category 2 is Data Acquisition. Fig 11 shows this conceptually as a collection of effects in the instrumentation system. However, at Pyestock the complete effect is always assessed by 'end-to-end' calibrations of the measuring instrument against the working standard. In the case of test frame load in Table 1, data acquisition errors are included in the last stage of Category 1 and the middle item of Category 4. (The important thing is that no item should be overlooked, but no item counted twice.)

Category 3 is Data Reduction. In Table 1, an elemental bias limit of 20 N is assessed due to nonlinearity of load calibration.

Category 4 in Fig 11 was introduced at Pyestock to cover important 'non-instrumentation' effects. In the case of test frame load, for example, these real effects are manifestations of test frame mechanics (stray friction, spring and gravity forces) which are found to be more important than any errors in the instrumentation. (The Bofors shear force load cell with its fundamental calibration, traceable to the National Standard, is responsible for very little error.)

When all the entries for elemental errors are completed for a particular basic measurement, such as test frame load in Table 1, these elemental bias limits are combined by root-sum-squares. In the case of Table 1 this amounts to 41 N.

Step 1 is complete when the separate elemental error tables for each of the basic measurements have been completed.

Step 2 (see Fig 10) consists of propagating the values of B_i from the basic measurements to the test result, R_k. It is convenient to do these calculations in separate tables such as Table 2, one for each flight condition. Values of B_i are entered in Table 2 and converted to %. The effect of each basic measurement is then propagated to each result, R_k, by means of the appropriate influence coefficient, thus

$$B_{k_i} = \theta'_{k_i} B_i \qquad (20)$$

For example in Table 2, for the result of R_k = SFCRD we have

i	x_i	B_i (%)	θ'_{Ki} (%/%)	B_{Ki} (%)
1	PSA	0.06	-0.12	-0.01
2	PA	0.25	-0.12	-0.03
etc				

Thence the root-sum-squares combination for all the inputs is

$$\pm B_K = \pm \sqrt{\sum_i B_{Ki}^2} \qquad (21)$$

for the result R_K = SFCRD , at Flight Condition 1. This process is depicted in bar chart form in Fig 9 for the result R_K = SFCRD for each of the 10 flight conditions.

7 THE CURVE SHIFT EFFECT

Engine performance results are always presented as graphical correlations against other suitable parameters. For example, SFC would be correlated against Net Thrust. Again, Net Thrust might be correlated against Pressure Ratio P7/P2 or, alternatively, against non-dimensional shaft speed, $NH/\sqrt{T1}$. Thus an isolated result of Net Thrust is of no value by itself without consideration of its correlating parameter.

The curve shift effect to be described in the present document is a phenomenon caused by bias errors (as defined at RAE(P)) propagated through the horizontal axis, z . Precision errors of z (as defined at RAE(P)) will have some effect on the scatter of points but, since this is likely to be small, it has been decided not to describe it in the present document, in order to concentrate attention on the more important effect of bias errors.

The following three Figures (14, 15, 16), taken from actual UETP results, illustrate the problem, while Fig 17 displays the theory to be explained in the text after it.

Fig 14 is an example of a comparison between the performance results from some UETP tests on a J57 engine measured in several international facilities, displayed as conventional graphs of Referred Net Thrust, y = FNRD, vs Pressure Ratio, z = P7/P2. Unfortunately, with this conventional presentation, the eye can hardly see the difference between facilities. However, a revised display in Fig 15 shows deviations ΔFNRD from a datum line, plotted against the same P7 P2, in which the differences between facilities can now be plainly seen. Fig 16 is

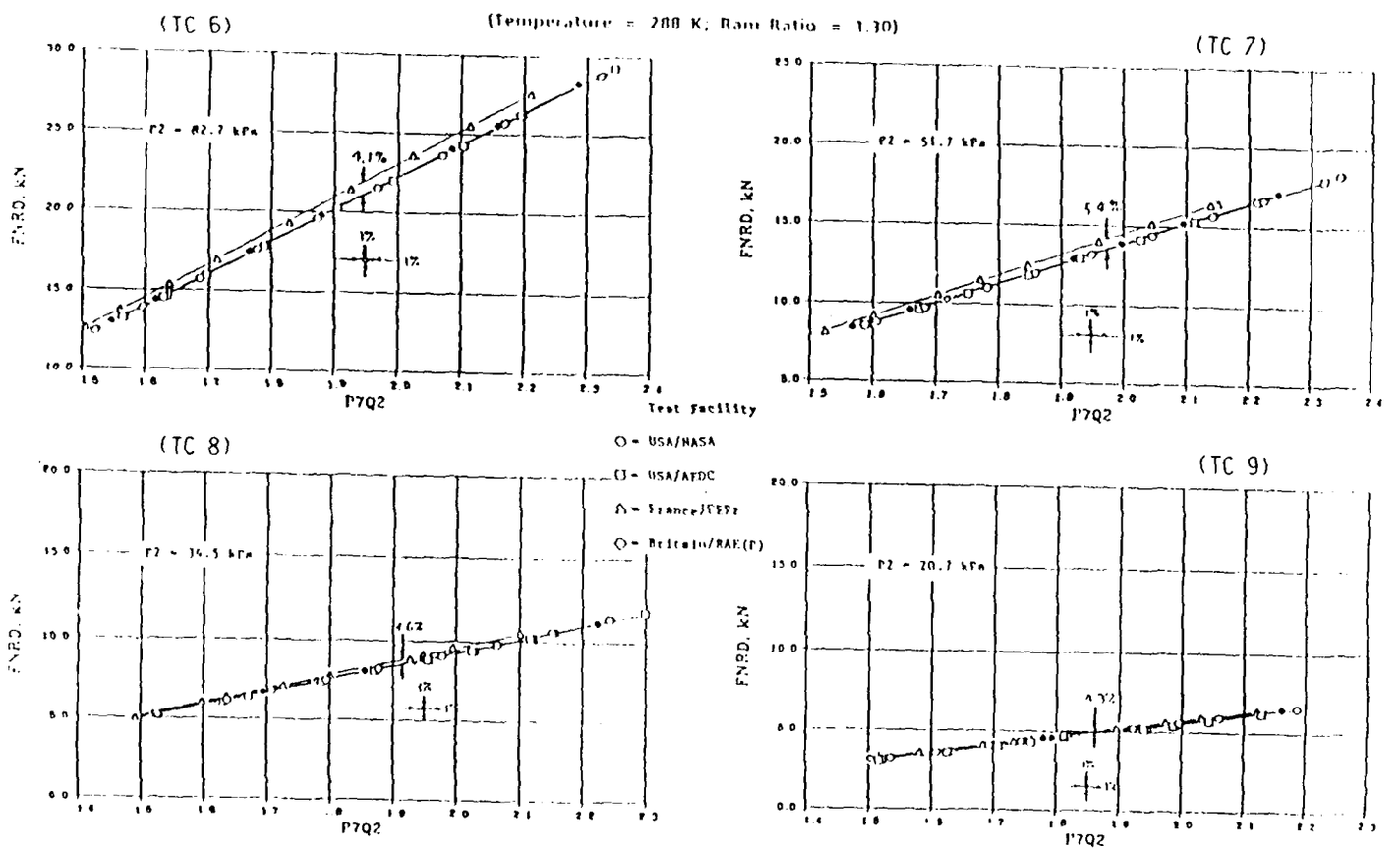
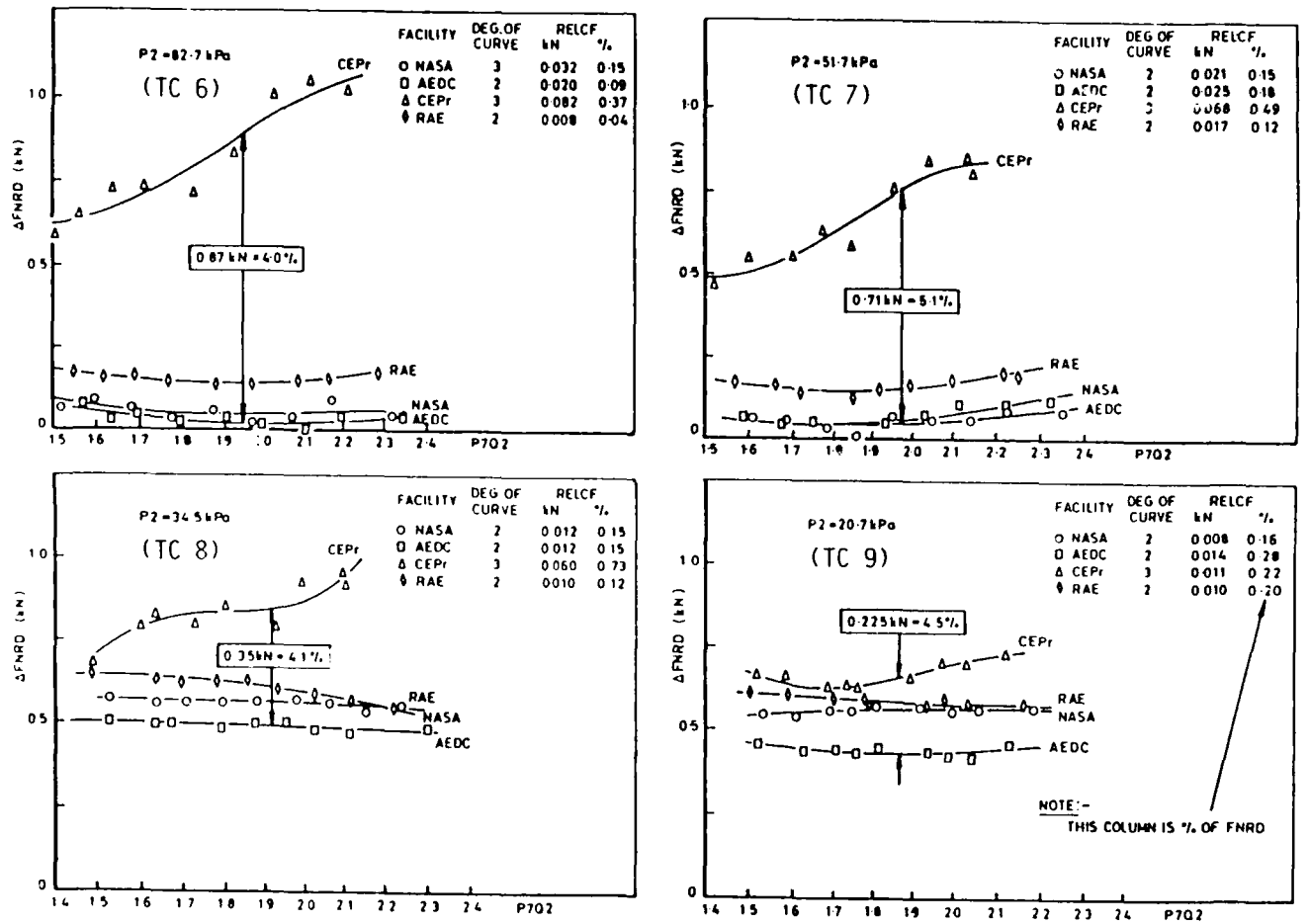


Fig 14 Example of conventional graphs of engine performance (FNRD vs P7/P2) (hard-to-see differences between test facilities)



ΔFNRD = DIFFERENCE FROM DATUM LINE)

Fig 15 Enhanced graphs of engine performance (ΔFNRD vs P7/P2) (easy-to-see differences between facilities)

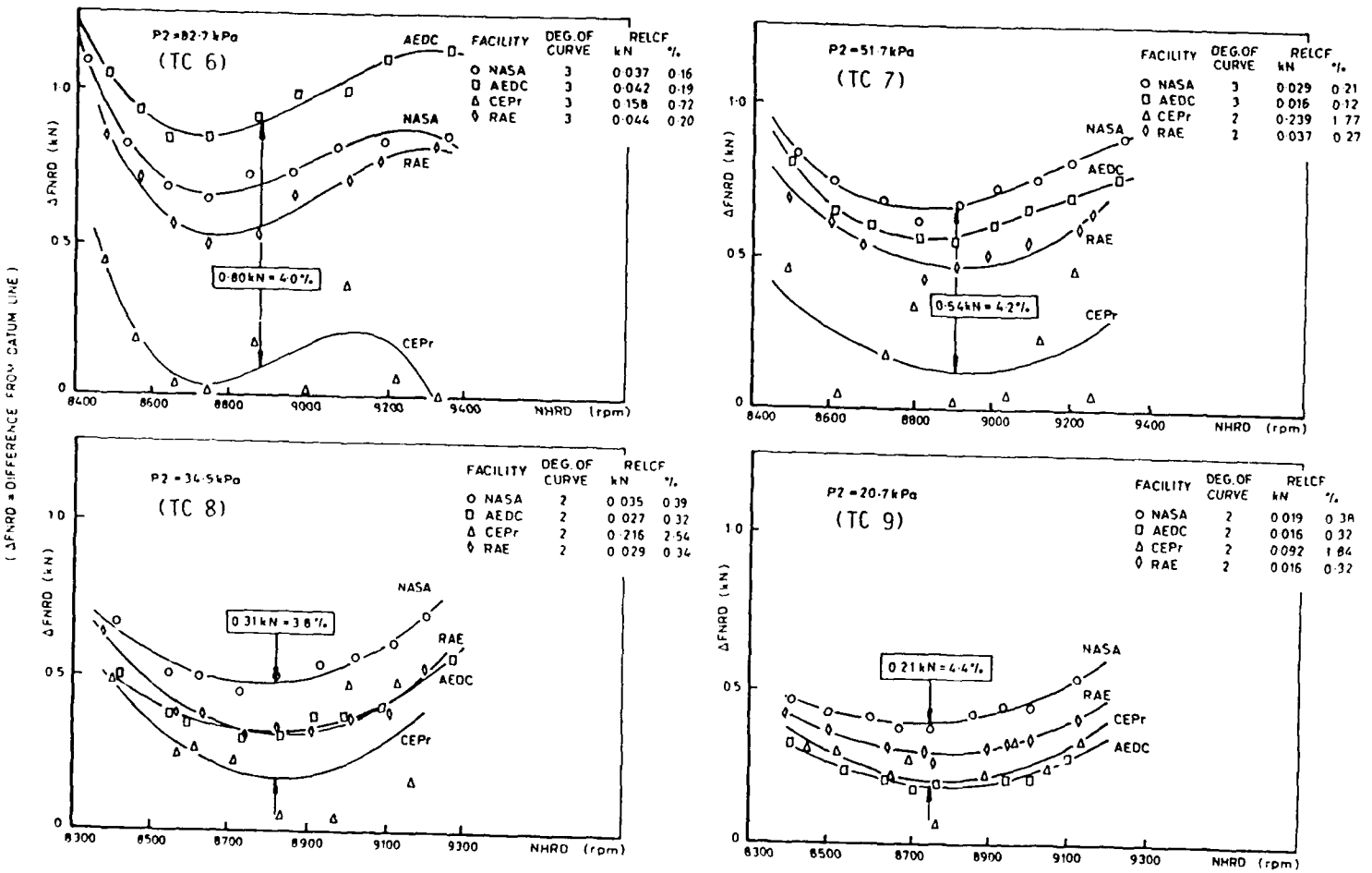


Fig 16 Alternative enhanced graphs of engine performance (ΔFNRD vs NHRD) (z = NHRD instead of z = P7/P2)

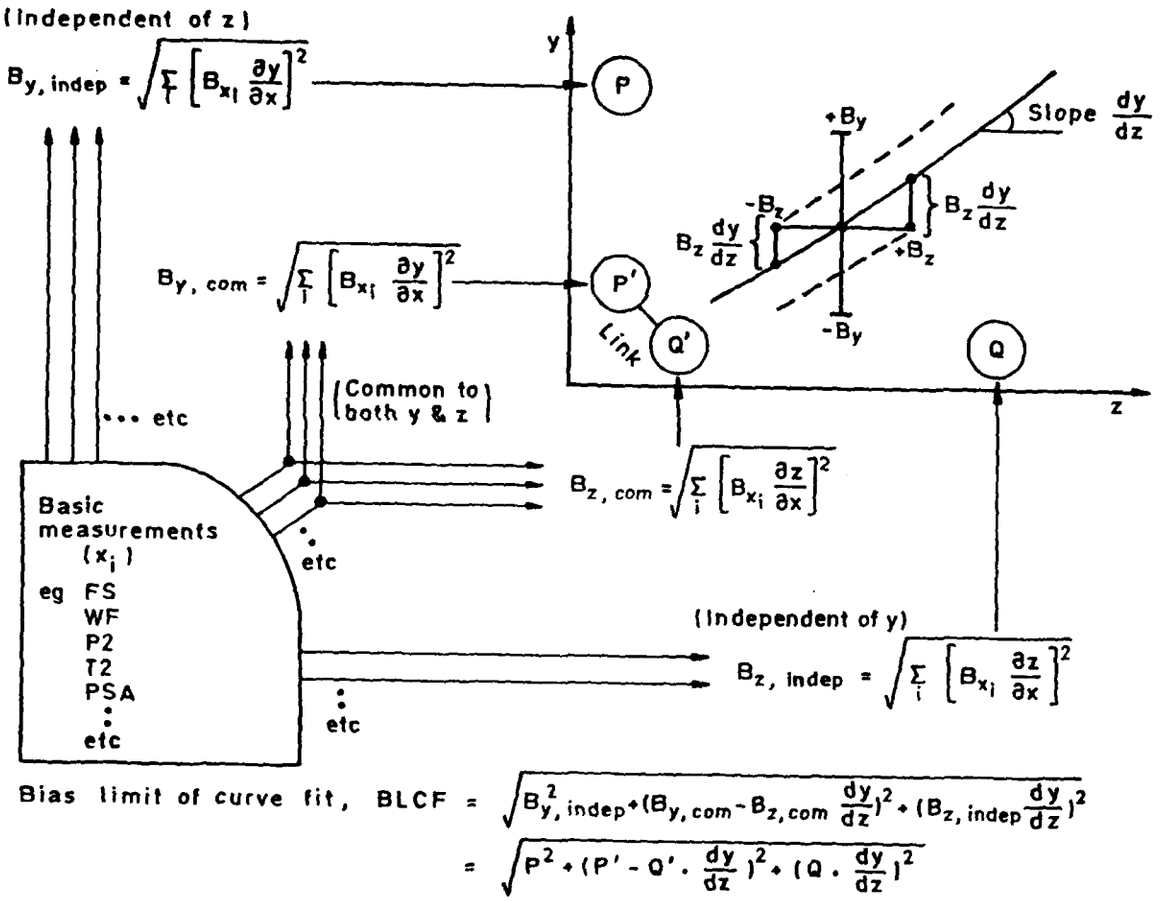


Fig 17 Graphical propagation of bias errors to performance curve

an alternative revised display, using z = NHRD instead of P7/P2. Comparing Figs 15 and 16, the pattern of the curves from the various facilities is seen to be drastically altered by changing the correlating parameter from z = P7/P2 to z = NHRD, even though the values of y = FNRD are exactly the same. So, it is the bias errors in P7/P2 (on the one hand) or the bias errors in NHRD (on the other hand) which make such a big difference to the performance curves of FNRD.

The mechanism of the graphical propagation of errors is presented in Fig 17, in terms of bias limits. The theory only applies to errors that remain constant in the performance curve. This is no problem with the DMP adopted by RAE(P) which regards such errors as bias. However, there would be a problem if some of these constant errors were regarded as precision, as was done by some UETP participants.

Errors in the Basic Measurements (x_i) can propagate to the curve in three ways, as depicted in Fig 17. The most obvious route [P] is travelled by errors that only affect 'y', and these limits are calculated by standard Abernethy methodology. The next route [Q] is travelled by errors that only affect 'z'. While these latter have no effect on the y-values of the data points, they do affect the y-position OF THE CURVE as indicated in Fig 17.

The third route [P]'.....[Q]' is followed by errors that affect 'y' and 'z' simultaneously. Their complete effect should be fully allowed for as indicated in Fig 17, BEFORE it is combined by root-sum-squares with the bias limits from route [P] and [Q]. This process yields the 'Bias Limit of Curve Fit', thus

$$BLCF = \sqrt{B_{y,indep}^2 + \left[B_{y,com} - B_{z,com} \cdot \frac{dy}{dz}\right]^2 + \left[B_{z,indep} \cdot \frac{dy}{dz}\right]^2} \quad . \qquad (22)$$

Equation (22) can often be simplified, according to circumstances. Thus, if the curve slope, dy/dz is small, it reduces simply to

$$BLCF = \sqrt{B_{y,indep}^2 + B_{y,com}^2} = B_{y,all} \quad , \qquad (23)$$

where $B_{y,all}$ is the total predicted bias limit of y, propagated from all basic measurements, using standard Abernethy methodology.

Again, if most of the error in 'z' are independent of 'y', it becomes

$$BLCF = \sqrt{B_{y,indep}^2 + \left[B_{z,indep} \cdot \frac{dy}{dz}\right]^2} \quad . \tag{24}$$

Or again, if most of the errors in 'z' are common to 'y', we get

$$BLCF = \left[B_{y,com} - B_{z,com} \cdot \frac{dy}{dz}\right] \quad . \tag{25}$$

An example of equation (24) is provided by the curves of FNRD vs P7Q2, where errors in 'y' and 'z' are practically all independent, and so the bias limits accumulate almost entirely by root-sum-squares as illustrated in the top of Fig 18.

The special case of equation (25) as applied to the SFC curves is illustrated in the bottom half of Fig 18 where the sense of the slope is seen to have a serious effect on BLCF. Errors in z = FNRD are obviously common to y = SFCRD = WFRD/FNRD via the variable FNRD, and some errors from other sources are common. Because of this special relationship, the positive limit of B_y is always associated with the negative limit of B_z, and vice versa. The consequence is that these bias limits will self-cancel when the slop is negative, but will build up when the slop is positive.

This self-cancelling effect due to the steep negative slope of the SFC curves in the UETP exercise at low P2 pressure levels, might explain why the interfacility spreads of the SFC curves were so surprisingly small at the difficult high altitude Test Condition 9.

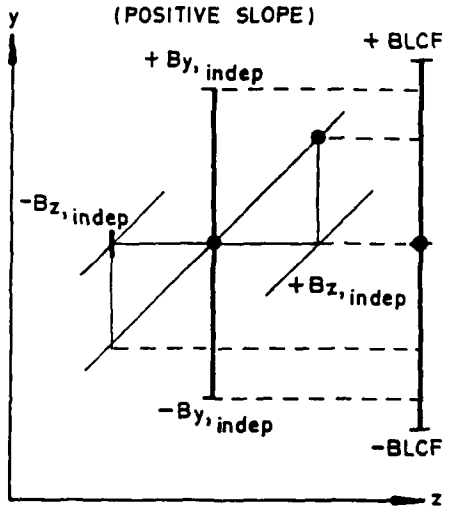
The last diagram at the bottom right of Fig 18, illustrates the build-up of the common errors in the graphs of Wa1RD vs NLRD due, in this case, to the positive slope. The common error here is the temperature, T1. The group WA1RD is related to $WA1\sqrt{T1}$, while NLRD is related to $NL/\sqrt{T1}$. Hence a positive error in T1 causes a positive $B_{z,com}$ and a negative $B_{y,com}$ which build up in a negative sense, as indicated in Fig 18. On the other hand, a negative error in T1 causes a negative $B_{z,com}$ and a positive $B_{y,com}$ which build up in a positive sense. It was suggested in Ref 11 that this important effect of common error in T1 could be a possible contributor to the so-called 'Cell Effect', of which the main symptom was an apparent difference on engine airflow between ground level test beds and altitude cells.

a) GENERAL CASE WITH INDEPENDENT ERRORS IN y AND z

e.g. FNRD V. P7Q2

BIAS LIMITS ACCUMULATE BY R.S.S.

NOTE:-
SEE FIG.17 FOR COMPLETE FORMULA FOR BLCF FOR ALL CASES.



b) SPECIAL CASES WITH COMMON ERROR IN y AND z

NEGATIVE SLOPE SFCRD V. FNRD
BIAS LIMITS SELF-CANCEL
(MOST ERRORS ARE COMMON)

POSITIVE SLOPE WA1RD V. NLRD
BIAS LIMITS BUILD UP
(T1 ERROR IS COMMON)

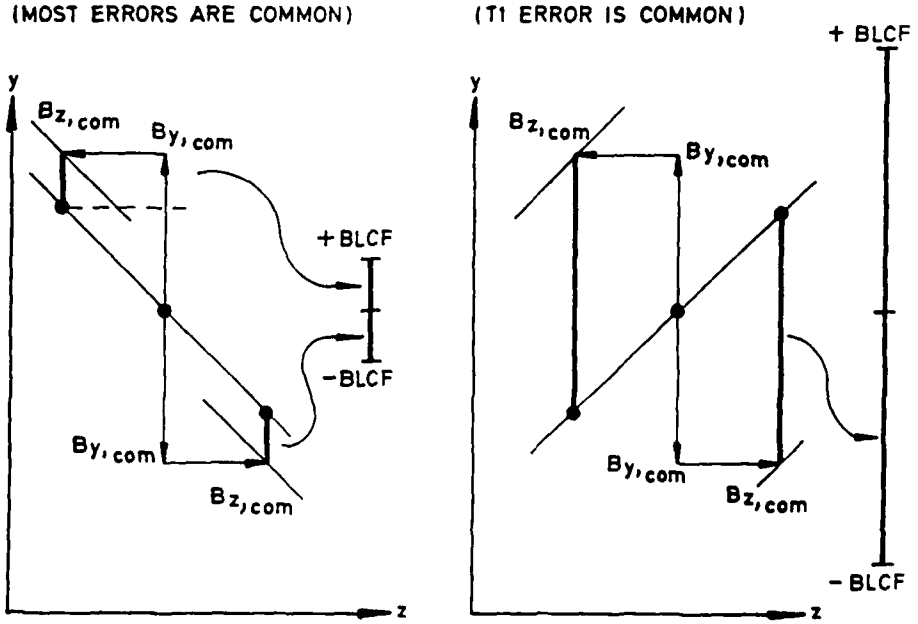


Fig 18 Illustration of extreme cases of curve shifts

8 HISTORICAL REVIEW OF THE METHODOLOGY

The methodology described in this Memorandum originated in the USA about 15 years ago[1]. At that time a rather different procedure, developed in the UK[6], was used at Pyestock. This latter took account of three classes of error:

(i) short-term random errors,

(ii) errors which are systematic in the short term, but random in the long term,

(iii) long term systematic errors.

At around the same time, contact was made with Abernethy on SAE E33 committee, where it was realised that both methodologies could be reconciled by the concept of 'the defined measurement process'. This allowed the 3-class MIDAP system to be simplified to the 2-class Abernethy system, and so the latter was introduced at Pyestock.

After several years experience here, culminating in the UETP exercise[2-4] in which a J57 engine was tested in several facilities on both sides of the Atlantic, the most valuable features have been consolidated. The procedures of 'prediction synthesis' are recognised as essential for assessing bias error limits, and the discipline of its formal framework helps to ensure that no source of error is overlooked, but none counted twice. A particularly useful feature is the separate 'elemental error table' (*eg* Table 1) for each basic measurement, or input parameter. These tables, together with the error propagation chart (*eg* Fig 9) are especially useful for identifying and tracking down the most important error sources. On the other hand, the effects of precision errors are considered to be best left to 'actual results analysis', without the need for the extensive effort involved in prediction synthesis.

Some aspects of the 'curve shift effect' had been studied at Pyestock more than a decade ago, with particular regard to T1 error[11]. But it was the stimulus of the UETP analysis which led to the full understanding of this phenomenon.

To sum up: we now have a good, practical uncertainty methodology which is understood and appreciated for engine testing in all NATO facilities.

Appendix A

EXPERIMENT WITH RANDOM NUMBERS TO DEMONSTRATE STATISTICAL PROPERTIES OF MEAN VALUES AND CURVE FITS

A.1 Objective of the experiment

When a curve is fitted to 'n' points of two-dimensional data by the method of least squares, it has to be recognised that the curve is only an estimate of where the true curve really is, as obscured by the random scatter of the data. The fitted curve is the best estimate, but the true curve might be anywhere within an uncertainty band whose limits can be calculated by analysis of the data in the form of the Random Error Limits of Curve Fit (± RELCF). Section A.3 demonstrates how well such RELCF calculations might cover a true curve when a sample of 20 points are analysed.

The statistical properties of a least squares curve are analogous to the properties of a mean value of a random sample. Since the latter is easier to explain and understand, this will be discussed first, in section A.2.

In both sections A.2 and A.3 the results are presented in alternative ways. The first and simplest way is to plot the theoretical limits about the true mean, or true curve, and note how well these limits enclose the experimental mean, or experimental curve. Unfortunately, in practice the true mean and true curve are usually not known and so it is more instructive to plot the experimental limits about the experimental mean, or experimental curve, and note how well these limits enclose the true mean, or true curve.

Thus the theory of the text books is put to the test with an actual sample of random numbers. In this experiment we know exactly what the true mean value and true curve really are, but we do not know in advance what values of random numbers will crop up. When the numbers do appear, it is interesting and instructive to see how the experimental means and curve fits fall within their estimated uncertainty bands.

A.2 Statistical properties of mean values

Random number generators will produce a series of numbers (R) from a rectangular distribution between 0 and 1, thus:

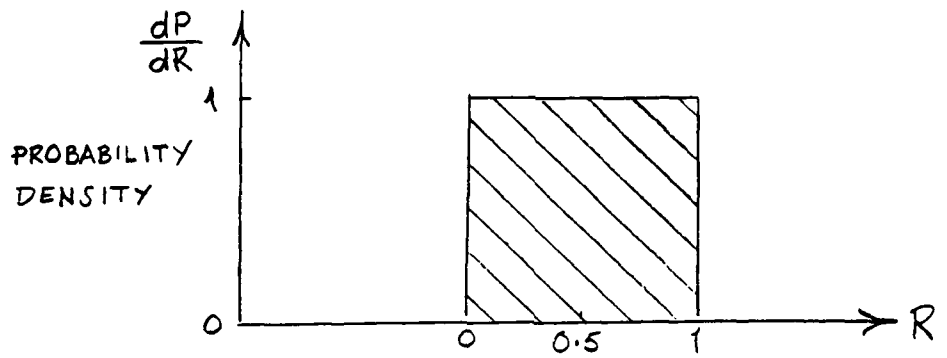


Fig A1 Rectangular distribution of random numbers

All values of R are equally likely between 0 and 1, with a true mean of 0.5. Note that the area under the curve equals 1 (*ie* total probability = 1).

To generate samples from a Gaussian distribution, we simply take a sequence of 12 numbers from the above, add them together and subtract the constant 6.00 from the sum:

$$G = \left(\sum_{i=1}^{12} R_i \right) - 6.00 .$$

It is shown in Ref A1 that the resulting number (G) will be a sample from a Gaussian distribution, with the true mean of 0 and the standard deviation of 1. Again, the area under the curve equals 1. (Note that $\sigma = 1$ only when 12 random numbers are taken from the generator.)

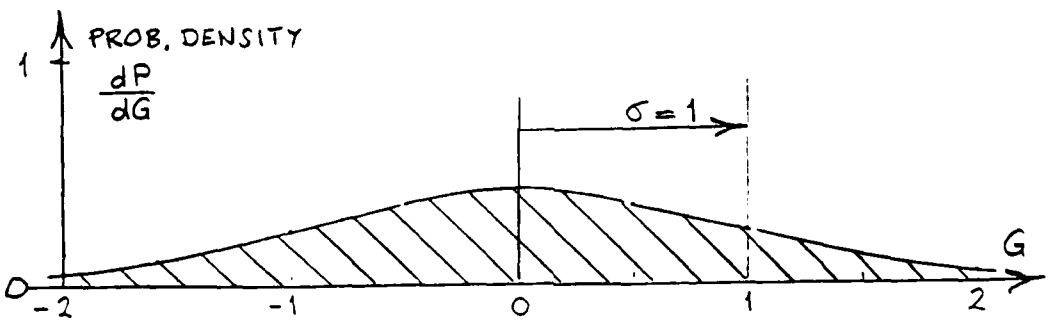


Fig A2 Gaussian distribution derived from 12 random numbers

For the present experiment, a series of 20 such values of G were taken and plotted out in Fig A3. After the first two values of G were available, their experimental mean was calculated:

$$\bar{G}_2 = \frac{(G_1 + G_2)}{2} .$$

Then after the next value, G_3 , arrived, the mean of 3 was calculated:

$$\bar{G}_3 = \frac{(G_1 + G_2 + G_3)}{3}$$

and so on up to $\bar{G}_{20}$.

At each stage, the experimental standard deviation was also calculated:

$$s = \sqrt{\frac{\Sigma(G - \bar{G})^2}{n - 1}} .$$

This is an estimate of the standard deviation, the true value of which we know to be exactly $\sigma = 1$.

In the top left of Fig A3 these experimental values of points, G and means, $\bar{G}$, are enclosed by the theoretical limits drawn about the true mean of zero. Here, the lines at $\pm 2\sigma$ are the 95% confidence limits of the points, G , while the $\pm 2\sigma$ limits of mean values are given by:

$$\sigma(\text{mean}) = \frac{\sigma(\text{points})}{\sqrt{n}} .$$

The $\pm 2\sigma$ limits remain a constant distance from the true mean, and it found that one out of 20 points falls outside these limits, as expected. The $\pm 2\sigma$(mean) limits begin quite wide apart where 'n' is small, but converge steadily as 'n' increases. All the experimental means are seen to lie inside these limits, while we might have expected one of them to fall outside, although the mean of 4 is quite near to the upper limit.

The top right of Fig A3 shows the theoretical probably distribution of the points, as already seen in the text above. The theoretical distribution of the

mean of 20 points is also shown. This is much tighter than the distribution of points, but the area under both curves equals 1 (*ie* total probability = 1).

Now, in a practical situation, the true mean (μ) and theoretical standard deviation (σ) would not be known. Instead, we must deal with the experimental means ($\bar{G}$) and experimental standard deviations (s). Hence the bottom left of Fig A3 shows the experimental 95% limits drawn about the experimental means. This time it is necessary to use a value of Student's "t_{95}" instead of the theoretical factor of 2. An interesting feature is that the $\pm ts$ limits, calculated just after the wide fourth point arrived, are much wider apart than the theoretical $\pm 2\sigma$ limits - *ie* with this sample of only $n = 4$, including the wide point, it seemed that a very scattered distribution was being sampled. However, by the time that $n = 20$ point had been taken, the $\pm ts$ limits had tightened such that 1 out of 20 points fell outside.

The experimental standard deviation of the mean is given by

$$s(\text{mean}) = \frac{s(\text{points})}{\sqrt{n}}$$

and the $\pm ts(\text{mean})$ limits, drawn about the experimental means, are seen to enclose the true mean easily.

A histogram of the experimental points is shown at the bottom right of Fig A3. This bears a rough resemblance to the theoretical distribution of points. Noting that 'relative frequency' is the same thing as 'probability', we expect the area of the histogram to be exactly 1 (as it is) just as for the area under the curves.

It is important to emphasise the distinction between the use of theoretical and experimental limits of the mean values. The former is a notional situation, useful for demonstration purposes where we can draw the $\pm 2\sigma(\text{mean})$ limits in advance and then watch the experimental means come along and fall within this band, as in the present experiment. The latter is the usual practical situation where the true mean is not known. This time we can only calculate $s(\text{mean})$ in retrospect, and expect the true mean to lie within the $\pm ts(\text{mean})$ band drawn about the experimental means.

A.3 Statistical properties of a curve fit

In principle, it would be possible to show how successive curve fits evolved as the same size (n) built up from $n = 4$ (which is the smallest size for a quadratic fit, with statistical calculations) to $n = 20$. This would

follow the procedure followed for the mean value results in section A.2 and Fig A3, but would be too complicated for this Appendix. Instead, let us suppose that 20 points of (z,y) data had been obtained, with a sample of Gaussian error the same as Fig A3 superimposed upon the true y values. The results of a quadratic curve fitted by least square to these data are shown in Fig A4. In the top left of the Figure, the theoretical $\pm 2\sigma$ limits (of individual points) and the $\pm 2\sigma$ limits (of the curve fit) are drawn about the true curve. As for the mean value example in Fig A3, it is seen that 1 out of 20 points falls outside these 2σ limits, as expected, while the curve fit values all lie within their $\pm 2\sigma$ limits - which is a little better than expected. Note that the $\pm 2\sigma$ limits of curve fit become wider apart as we move towards the ends of the curve.

As before, in a practical situation the true curve and true standard deviation would not be known. And so it is necessary to draw the experimental limits about the experimental curve fit. The theoretical standard deviation, σ, is replaced by the experimental RSD and, instead of $\pm 2\sigma$ (of curve fit), we plot $\pm t s$ (of curve fit), *ie* the Random Error Limits of Curve Fit, $\pm$RELCF. To plot the 95% confidence limits of individual points, the limits $\pm t.RSD$ should strictly be combined with RELCF by root-sum-squares, to give the Random Error Limits of Individual Points:

$$\pm RELIP = \pm [(t.RSD)^2 + RELCF^2]^{\frac{1}{2}} .$$

As the number of points, n, increases, the value of RELCF decreases, while t.RSD stabilises. In fact, with the value of n as high as 20, it is a good approximation to say

$$\pm RELIP \simeq \pm t.RSD$$

and these limits are plotted about the curve fit in the bottom half of Fig A4. It is seen that 1 out of 20 points falls outside the $\pm t.RSD$ limits, while the curve fit values all lie within $\pm$RELCF.

In the top right of Fig A4 are to be seen the smooth theoretical Gaussian probability distributions, centered on the true curve. The distribution of points is exactly the same as was shown in the top right of Fig A3 but, instead of the distribution of mean values, we now have the distribution of curve fits, evaluated at two places - *ie* near the middle of the curve where it is tightest, and at the ends of the curve where it is widest.

The histogram of the experimental distribution of points is shown at the bottom right of Fig A4, centred on the curve fit. This is the same as the histogram in Fig A3, centred on the experimental mean of 20 points, with the area of the histogram equal to 1 exactly (*ie* total probability = 1).

A.4 Concluding remarks

This experiment has analysed a sample of 20 random numbers from a Gaussian distribution. At the beginning, the writer did not know what the numbers would be and, if the experiment were to be repeated, a different of numbers would appear. In fact, the values in Figs A3 and A4 should never occur again. However, these numbers do demonstrate the sort of way that mean values and curve fits are expected to occur within the limits that can be calculated from the data by statistical principles.

Finally, it should be remembered that this Appendix only deals with the properties of random errors from a Gaussian distribution - sometimes called 'precision errors'. If any systematic (or bias) errors exist, then their effect would be superimposed on top of the precision errors.

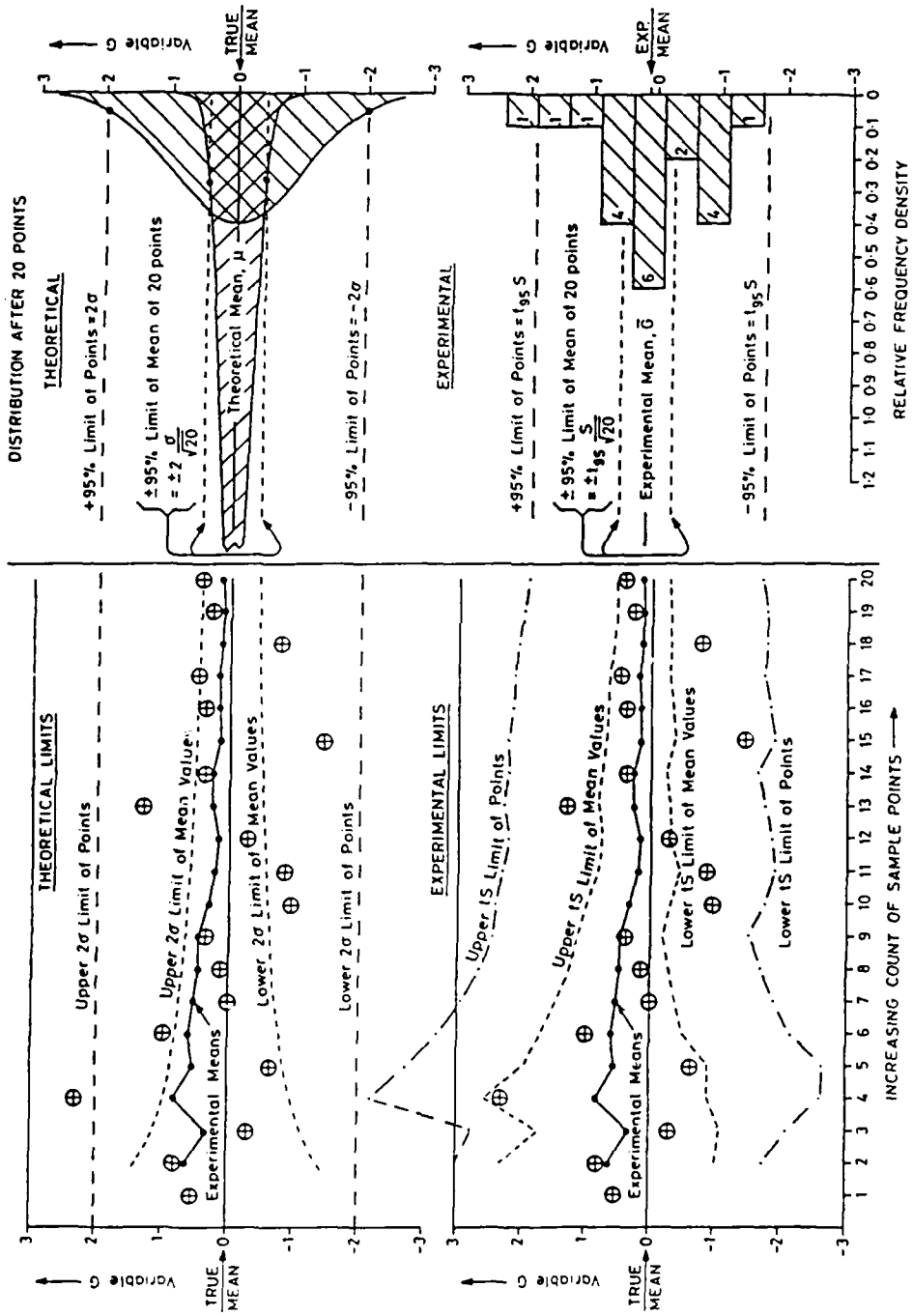


Fig A3 Statistics of an experimental sample from a Gaussian distribution

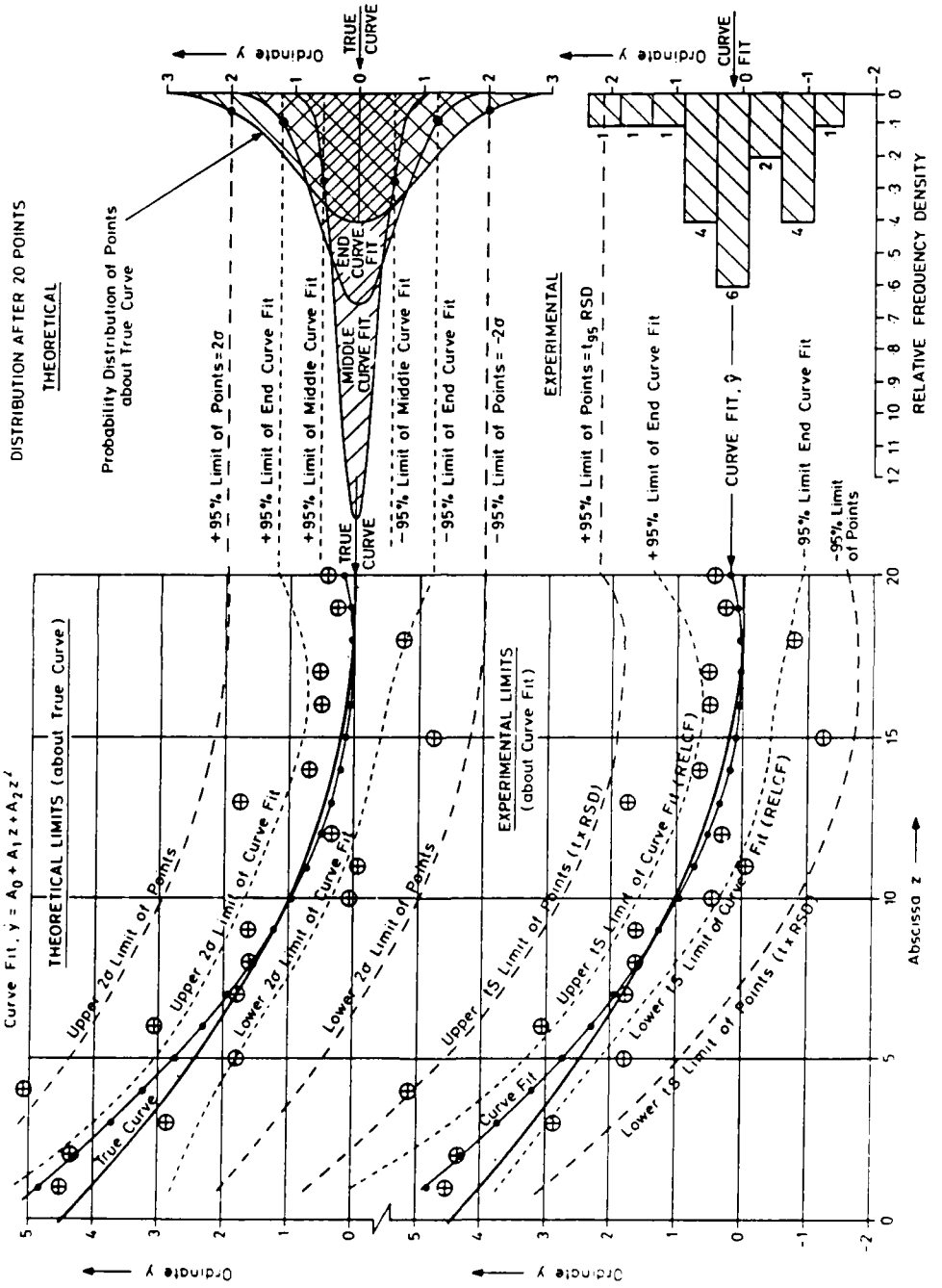


Fig A4 Curve fitting of an experimental sample from a Gaussian distribution

NOTATION USED IN APPENDIX A

Symbol	Description
A_0, A_1, A_2	coefficients of least squares curve fit
G	Gaussian random number
n	number of values in a sample
P	probability, *ie* relative frequency of an occurrence
R	rectangular random number
RELCF	random error limit of curve fit
RELIP	random error limit of individual points
RSD	residual standard deviation
s	experimental standard deviation
t	Student's statistic (at 95% probability in this Appendix)
y	ordinate value of a two-dimensional point (vertical axis)
$\hat{y}$	curve fit value of y
z	abscissa value of a two-dimensional point (horizontal axis)
σ	theoretical standard deviation (true value)
μ	theoretical mean value (true value)

REFERENCE IN APPENDIX A

No.	Author	Title, etc
A1	R.W. Hamming	Numerical methods for scientists and engineers. McGraw Hill, 2nd edition (1973)

Appendix B

OUTLIER DETECTION

B.1 Outliers from a mean value

Two methods can be recommended, (a) due to Thompson[B1] and, (b) due to Grubbs[B2]. The former is more active in that it will catch more real wild points, but it is more prone to reject good points. The latter will reject few good points, but will not catch as many bad points. These accept/reject properties were confirmed by Abernethy using Monte Carlo simulation, giving the results shown in the following two figures. Hence Grubbs' method can safely be left to automatic computer operation, but Thompson's method should only be used to 'flag' possible outliers for further investigation.

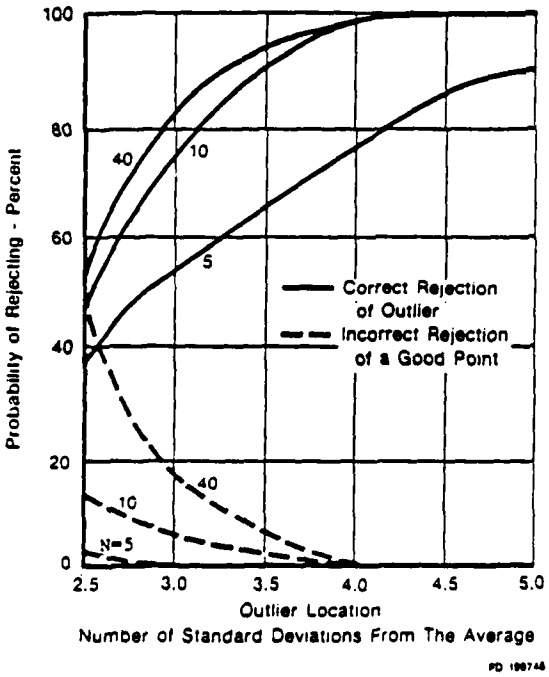


Fig B1 α,β error in Thompson outlier test (based on one outlier in each of 100 samples of sizes 5, 10 and 40)

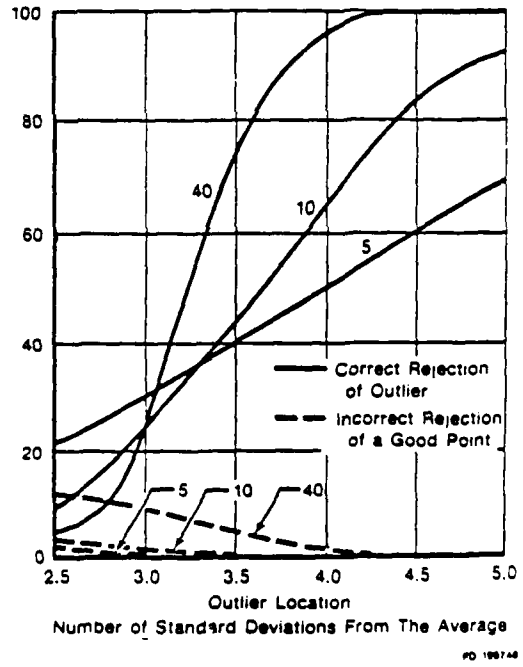


Fig B2 α,β error in Grubbs outlier test (based on one outlier in each of 100 samples of sizes 5, 10 and 40)

B.1.1 Thompson's Tau Test

From a sample of n values of x, calculate the mean value $\bar{x}$ and the sample standard deviation

$$s^* = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}} . \qquad \text{(B-1)}$$

Note that the divisor here is n, rather than $(n - 1)$. For any suspected outlier, x_j calculate

$$\tau_j = \frac{|x_j - \bar{x}|}{s^*} . \qquad \text{(B-2)}$$

The point is an outlier if τ_j is greater than the critical value in Table 1.

As an example, take the $n = 20$ points of data in Fig A3 of Appendix A, where we might suspect that point number 4 is an outlier. Thus, (using the symbol 'x' instead of 'G')

$$x_4 = 2.307$$

$$\bar{x} = 0.148$$

$$s^* = 0.844$$

$$\tau_4 = \frac{|2.307 - 0.148|}{0.844} = 2.558 .$$

From Table B1, the critical value is 1.934 for $n = 20$ at the 5% level of significance, hence this point would be flagged as a possible outlier. It is interesting to note that, in fact, ALL the points in Fig A3 are genuine values from a Gaussian distribution and so, in this example, Thompson's Tau Test has been overactive in rejecting a good point! This illustrates the sound advice that this test should only be used to flag suspect points for further investigation.

B.1.2 Grubbs' T Test

From a sample of x data, calculate the mean value $\bar{x}$ and the standard deviation

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}} . \qquad \text{(B-3)}$$

Note that this time the divisor is the usual (n - 1) rather than n . For any suspected outlier, x_j , calculate

$$T_j = \frac{|x_j - \bar{x}|}{s} . \qquad \text{(B-4)}$$

The point is an outlier if T_j is greater than the critical value in Table 2.

Taking the same example from the data in Fig A3, with point number 4 suspected as an outlier, we have

$$x_4 = 2.307$$

$$\bar{x} = 0.148$$

$$s = 0.866$$

$$T = \frac{|2.307 - 0.148|}{0.866} = 2.493 .$$

From Table B2, the critical value is 2.56 for $n = 20$ at the 5% confidence level, hence the point is correctly NOT classified as an outlier. Thus Grubbs' test could safely have been left to automatic computer operation.

B.2 Outliers from a curve fit

Outlier detection methods are not so well established for curve fits as for mean values. However, there will usually be a computer VDU graph available for inspection, such as Fig 7 in the main text, to help the judgement. The current practice at Pyestock for examining such scatter is explained first, in section B.2.1. An alternative method for possible future use is discussed in section B.2.2.

B.2.1 Current practice at Pyestock for checking outliers from curve fits

The computer program which produces the curve fits such as Fig 7 of the main text, will draw lines of ±RELCF and ±RELIP on each side of the curve, where

$$\pm \text{RELIP} = \pm \left[(t_{95} s_r)^2 + \text{RELCF}^2 \right]^{\frac{1}{2}} , \qquad \text{(B-5)}$$

ie

$$\pm t_{95} s_{IP} = \pm \left[(t_{95} s_r)^2 + (t_{95} s_{\hat{y}})^2 \right]^{\frac{1}{2}} , \qquad \text{(B-6)}$$

where $s_r = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n - p - 1}}$, is the residual standard deviation

p is the highest power of the polynomial

and $s_{\hat{y}}$ is the standard deviation of the curve fit.

Dealing first with a straight line, we would have

$$s_{\hat{y}} = s_r \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]^{\frac{1}{2}} . \qquad \text{(B-7)}$$

Removing the t_{95} factor from equation (B-6)

$$s_{IP} = \left[s_r^2 + s_{\hat{y}}^2 \right]^{\frac{1}{2}} \qquad \text{(B-8)}$$

$$= s_r \left[1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]^{\frac{1}{2}} . \qquad \text{(B-9)}$$

(Equation (B-9) agrees with a standard formula in Ref B3.)

With the current outlier detection practice at Pyestock, we would expect 1 in 20 points to fall just outside the ±RELIP boundaries. For example, point number 4 in the curve fit shown in Fig A4 of Appendix A would be accepted as a good point. Any point which fell much further away from the curve would be suspected as an outlier and hence a candidate for deletion. If deleted, the curve fit would be recalculated with one less data point, or it might be possible to take another point to keep the numbers up.

B.2.2 A possible alternative outlier test for curve fits

Although the current practice for outlier detection at Pyestock is as described in section B.2.1, various alternative methods are becoming available. A good candidate for possible future use, due to Tietjens et al[B4], was one of a collection recommended by Abernethy and Ringhiser[B5]. This is now described, with a comment on a tricky piece of philosophy that it contains. Tietjens considers the theoretical model for a point (x_i, y_i) deviating from a straight line

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \,, \qquad \text{(B-10)}$$

where β_1 and β_2 are 'true' coefficients and ε is a Gaussian error. The least squares estimate of the line as shown in Fig B3 is

$$\hat{y}_i = b_1 + b_2 x_i \,. \qquad \text{(B-11)}$$

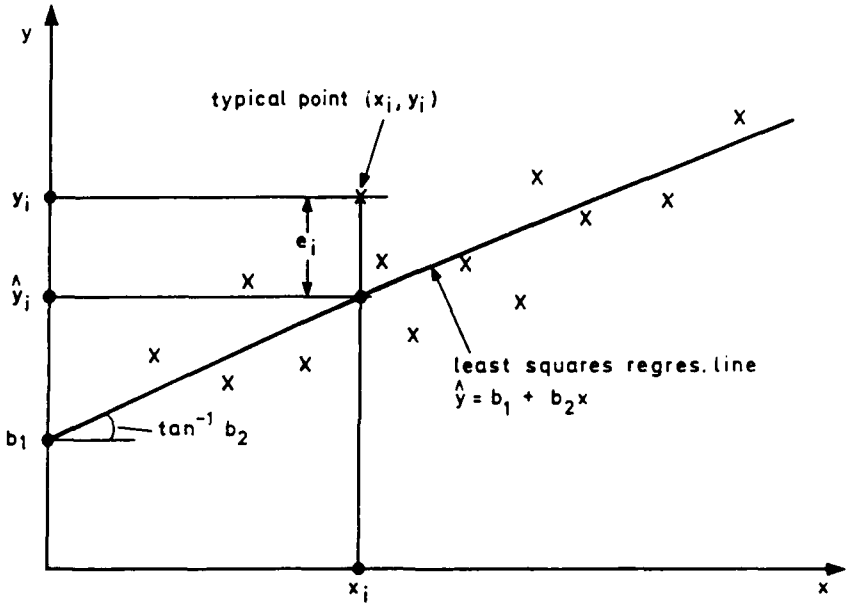


Fig B3 Residual deviations from regression line

The observed residuals (*ie* deviations from the line) are

$$e_i = y_i - \hat{y}_i \qquad \text{(B-12)}$$

$$= y_i - (b_1 + b_2 x_i) \qquad \text{(B-13)}$$

and the residual standard deviation is

$$\hat{\sigma}_r = \sqrt{\frac{\Sigma e^2}{n - 2}} \,. \qquad \text{(B-14)}$$

Note that there are several alternative symbols for this item

$$\hat{\sigma}_r = s_r = \text{RSD} .$$

Tietjens states that the estimated standard deviation of e is

$$s_i = \hat{\sigma}_r \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]^{1/2} . \qquad \text{(B-15)}$$

This bears an uncanny resemblance to equation (B-9) for the current Pyestock method, except for the minus signs instead of plus signs! A possible explanation is as follows.

Returning to equation (B-12)

$$e_i = y_i - \hat{y}_i$$

and applying the rule: 'the variance of a difference is equal to the SUM of the variances', we get

$$\text{var(of } e_i) = \text{var(of } y_i) + \text{var(of } \hat{y}_i) \qquad \text{(B-16)}$$

ie

$$s_r^2 = s_i^2 + s_{\hat{y}_i}^2 \qquad \text{(B-17)}$$

therefore

$$s_i^2 = s_r^2 - s_{\hat{y}_i}^2 \qquad \text{(B-18)}$$

$$= s_r^2 - s_r^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \qquad \text{(B-19)}$$

$$= s_r^2 \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \qquad \text{(B-20)}$$

therefore

$$s_i = s_r \left[1 - \frac{1}{n} - \frac{(x - \bar{x})^2}{\sum (x - \bar{x})^2} \right]^{1/2} , \qquad \text{(B-21)}$$

which is the same as Tietjens' equation, equation (B-15).

So, it would seem that Tietjens takes s_i to be the 'pure' standard deviation of the residual errors, whereas s_{IP} in equation (B-9) of the Pyestock method is the predicted standard deviation of the scattered points as augmented by the uncertainty of the curve fit.

To proceed with Tietjens, the ratio of the maximum deviation to its standard deviation is found

$$R = \max \left| \frac{e_i}{s_i} \right| . \qquad \text{(B-22)}$$

This is compared with critical values in Tables B3. If R exceeds the critical value, that point is declared to be an outlier, deleted, and then the curve fit is recalculated.

The above theory can be extended to higher orders of polynomial curve fits. Thus from equation (B-18)

$$s_i = \left[s_r^2 - s_{\hat{y}}^2 \right]^{\frac{1}{2}} \qquad \text{(B-23)}$$

$$= \left[RSD^2 - \left(\frac{RELCF}{t_{95}} \right)^{\frac{1}{2}} \right] . \qquad \text{(B-24)}$$

All the values for equation (24) are available from the Pyestock computer program for polynomial curves up to the cubic.

The ratio for the maximum deviation

$$R = \max \left| \frac{e_i}{s_i} \right| . \qquad \text{(B-22) bis}$$

is then compared with critical values in Tables B4, taken from Ref B6.

This procedure is illustrated by the example of the quadratic curve fit in Fig A4 of Appendix A. Of the 20 points scattered about the curve, point number 4 is a suspect outlier. For this point we have

$$e_i = 2.3$$

$$s_i = \left[0.95^2 - \left(\frac{0.65}{2.11}\right)^2 \right]^{\frac{1}{2}} = 0.90$$

therefore $$R = \frac{2.3}{0.9} = 2.55 .$$

From Table B4 the critical values are

α	0.10	0.05	0.01
R	2.62	2.77	3.06

and so the suspect point number 4 is declared 'not guilty', even at the 10% level.

This verdict agrees with that of the current Pyestock method which, by inspection of points on Fig A4, would accept 1 point in 20 lying just outside the $\pm t_{95}s$ limits of points (also called $\pm$RELIP).

Tables B1 and B2

TESTING FOR OUTLIERS FROM MEAN VALUES

(Extracted from Ref B1)

Table 1

Thompson's Tau

(Extracted from Ref B2)

Table 2

Rejection values for Grubbs' Method

Table 1. Thompson's Tau

Sample Size N	Level of Significance P=10%	5%	2%	1%
3	1.3968	1.4099	1.41352	1.414039
4	1.559	1.6080	1.6974	1.7147
5	1.611	1.757	1.969	1.9175
6	1.631	1.814	1.973	2.0509
7	1.640	1.848	2.040	2.142
8	1.644	1.870	2.087	2.207
9	1.647	1.885	2.121	2.256
10	1.648	1.896	2.146	2.294
11	1.648	1.904	2.166	2.324
12	1.649	1.910	2.183	2.348
13	1.649	1.915	2.196	2.368
14	1.649	1.919	2.207	2.385
15	1.649	1.923	2.216	2.399
16	1.649	1.926	2.224	2.411
17	1.649	1.928	2.231	2.422
18	1.649	1.931	2.237	2.432
19	1.649	1.932	2.242	2.440
20	1.649	1.934	2.247	2.447
21	1.649	1.936	2.251	2.454
22	1.649	1.937	2.255	2.460
23	1.649	1.938	2.259	2.465
24	1.649	1.940	2.262	2.470
25	1.649	1.941	2.264	2.475
26	1.648	1.942	2.267	2.479
27	1.648	1.942	2.269	2.483
28	1.648	1.943	2.272	2.487
29	1.648	1.944	2.274	2.490
30	1.648	1.944	2.275	2.493
31	1.648	1.945	2.277	2.495
32	1.648	1.945	2.279	2.498
∞	1.64485	1.95996	2.32634	2.57582

Table 2. Rejection Values for Grubbs' Method

Sample Size N	Level of Significance P = 5%	P = 2.5%	P = 1%
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82
19	2.53	2.68	2.85
20	2.56	2.71	2.88
21	2.58	2.73	2.91
22	2.60	2.76	2.94
23	2.62	2.78	2.96
24	2.64	2.80	2.99
25	2.66	2.82	3.01
30	2.75	2.91	
35	2.82	2.98	
40	2.87	3.04	
45	2.92	3.09	
50	2.96	3.13	
60	3.03	3.20	
70	3.09	3.26	
80	3.14	3.31	
90	3.18	3.35	
100	3.21	3.38	

Table B3

TESTING FOR A SINGLE OUTLIER IN SIMPLE LINEAR REGRESSION

(Extracted from Ref B4)

Critical Values of R_n for Detecting One Outlier in Simple Linear Regression

Sample Size (n)	α = .10	α = .05	α = .01
4	1.41	1.41	1.41
5	1.69	1.71	1.73
6	1.88	1.92	1.97
7	2.01	2.07	2.16
8	2.10	2.19	2.31
9	2.18	2.28	2.43
10	2.24	2.35	2.53
11	2.30	2.43	2.64
12	2.35	2.48	2.70
14	2.43	2.57	2.80
16	2.50	2.64	2.92
18	2.56	2.71	2.99
20	2.60	2.76	3.06
24	2.69	2.85	3.17
30	2.79	2.97	3.28
36	2.86	3.03	3.35
48	2.97	3.15	3.51
60	3.04	3.21	3.50
100	3.22	3.40	3.75

Table B4

OUTLIER TESTING IN POLYNOMIAL REGRESSION

(Extracted from Ref B6)

TABLE 1—*Upper Bound for Critical Values for Studentized Residual*

(α = .10)

n	q									
	1	2	3	4	5	6	8	10	15	25
5	1.87									
6	2.00	1.89								
7	2.10	2.02	1.96							
8	2.18	2.12	2.03	1.91						
9	2.24	2.20	2.13	2.05	1.92					
10	2.30	2.26	2.21	2.15	2.06	1.92				
12	2.39	2.37	2.33	2.29	2.24	2.17	1.93			
14	2.47	2.45	2.42	2.39	2.36	2.32	2.19	1.94		
16	2.53	2.51	2.50	2.47	2.45	2.42	2.30	2.20		
18	2.58	2.57	2.56	2.54	2.52	2.50	2.44	2.35		
20	2.63	2.62	2.61	2.59	2.58	2.56	2.52	2.46	2.11	
25	2.72	2.72	2.71	2.70	2.69	2.68	2.66	2.63	2.50	
30	2.80	2.79	2.79	2.78	2.77	2.77	2.75	2.73	2.66	2.13
35	2.86	2.85	2.85	2.85	2.84	2.84	2.82	2.81	2.77	2.55
40	2.91	2.91	2.90	2.90	2.90	2.89	2.88	2.87	2.84	2.72
45	2.95	2.95	2.95	2.95	2.94	2.94	2.93	2.93	2.90	2.82
50	2.99	2.99	2.99	2.99	2.98	2.98	2.98	2.97	2.95	2.89
60	3.06	3.06	3.05	3.05	3.05	3.05	3.05	3.04	3.03	3.00
70	3.11	3.11	3.11	3.11	3.11	3.11	3.10	3.10	3.09	3.07
80	3.16	3.16	3.16	3.15	3.15	3.15	3.15	3.15	3.14	3.12
90	3.20	3.20	3.19	3.19	3.19	3.19	3.19	3.19	3.18	3.17
100	3.23	3.23	3.23	3.23	3.23	3.23	3.23	3.22	3.22	3.21

Table B4 (concluded)

(α = .05)

n	q									
	1	2	3	4	5	6	8	10	15	25
5	1.92									
6	2.07	1.93								
7	2.19	2.08	1.94							
8	2.28	2.20	2.10	1.94						
9	2.35	2.29	2.21	2.10	1.95					
10	2.42	2.37	2.31	2.22	2.11	1.95				
12	2.52	2.49	2.45	2.39	2.33	2.24	1.96			
14	2.61	2.58	2.55	2.51	2.47	2.41	2.25	1.96		
16	2.68	2.66	2.63	2.60	2.57	2.53	2.43	2.26		
18	2.73	2.72	2.70	2.68	2.65	2.62	2.55	2.44		
20	2.78	2.77	2.76	2.74	2.72	2.70	2.64	2.57	2.15	
25	2.89	2.88	2.87	2.86	2.84	2.83	2.80	2.76	2.60	
30	2.96	2.96	2.95	2.94	2.93	2.93	2.93	2.88	2.79	2.17
35	3.03	3.02	3.02	3.01	3.00	3.00	2.98	2.97	2.91	2.64
40	3.08	3.08	3.07	3.07	3.06	3.06	3.05	3.03	3.00	2.84
45	3.13	3.12	3.12	3.12	3.11	3.11	3.10	3.09	3.06	2.96
50	3.17	3.16	3.16	3.16	3.15	3.15	3.14	3.14	3.11	3.04
60	3.23	3.23	3.23	3.23	3.22	3.22	3.22	3.21	3.20	3.15
70	3.29	3.29	3.28	3.28	3.28	3.28	3.27	3.27	3.26	3.23
80	3.33	3.33	3.33	3.33	3.33	3.33	3.32	3.32	3.31	3.29
90	3.37	3.37	3.37	3.37	3.37	3.37	3.36	3.36	3.36	3.34
100	3.41	3.41	3.40	3.40	3.40	3.40	3.40	3.40	3.39	3.38

(α = .01)

n	q									
	1	2	3	4	5	6	8	10	15	25
5	1.98									
6	2.17	1.98								
7	2.32	2.17	1.98							
8	2.44	2.32	2.18	1.98						
9	2.54	2.44	2.33	2.18	1.99					
10	2.62	2.55	2.45	2.33	2.18	1.99				
12	2.76	2.70	2.64	2.56	2.46	2.34	1.99			
14	2.86	2.82	2.78	2.72	2.65	2.57	2.35	1.99		
16	2.95	2.92	2.88	2.84	2.79	2.73	2.53	2.35		
18	3.02	3.00	2.97	2.94	2.90	2.85	2.75	2.59		
20	3.08	3.06	3.04	3.01	2.98	2.95	2.87	2.76	2.20	
25	3.21	3.19	3.18	3.16	3.14	3.12	3.07	3.01	2.73	
30	3.30	3.29	3.28	3.26	3.25	3.24	3.21	3.17	3.04	2.21
35	3.37	3.36	3.35	3.34	3.34	3.33	3.30	3.28	3.19	2.81
40	3.43	3.42	3.42	3.41	3.40	3.40	3.38	3.36	3.30	3.09
45	3.48	3.47	3.47	3.46	3.46	3.45	3.44	3.43	3.38	3.23
50	3.52	3.52	3.51	3.51	3.51	3.50	3.49	3.48	3.45	3.34
60	3.60	3.59	3.59	3.59	3.58	3.58	3.57	3.56	3.54	3.48
70	3.65	3.65	3.65	3.65	3.64	3.64	3.64	3.63	3.61	3.57
80	3.70	3.70	3.70	3.70	3.69	3.69	3.69	3.64	3.67	3.64
90	3.74	3.74	3.74	3.74	3.74	3.74	3.73	3.73	3.72	3.70
100	3.78	3.78	3.78	3.77	3.77	3.77	3.77	3.77	3.76	3.74

n = number of observations

q = number of independent variables (including count for intercept if fitted)

NOTATION USED IN APPENDIX B

Roman symbols	Description
b_1	first coefficient in linear regression
b_2	second coefficient in linear regression
e	vertical deviation of point from regression line
n	number of points
p	highest power in a polynomial equation
q	number of coefficients (including the intercept) in a polynomial equation (Table 4)
R	max ratio of a residual to its standard deviation
RELCF	random error limit of curve fit
RELIP	random error limit of individual point
RSD	residual standard deviation
s	Grubbs' standard deviation, with (n - 1) divisor
s*	Thompson's standard deviation, with n divisor
s_i	'pure' standard deviation of residuals, used by Tietjens
s_{IP}	standard deviation of individual point
s_r	residual standard deviation
$s_{\hat{y}}$	standard deviation of curve fit
t_{95}	Student's statistic at 95% confidence
T	Grubb's statistic
x	either: a one-dimensional random variable or: a value of two-dimensional data on the horizontal axis
x_i	an individual value of x
x_j	a suspected outlier of x-data
$\bar{x}$	mean value of x-data
y	a value of two-dimensional data on the vertical axis
y_i	an individual value of y
$\hat{y}_i$	a curve-fit value of y

Greek symbols	
α	either: type of error due to incorrect rejection or: level of significance in critical tables
β	type of error due to incorrect acceptance
β_1	first regression coefficient in theoretical model
β_2	second regression coefficient in theoretical model
ϵ	theoretical Gaussian error in regression model
$\hat{\sigma}_r$	residual standard deviation
τ	Thompson's tau statistic

Notes:

1 Much of the notation in this Appendix is that of the original references.

2 The symbols $\hat{\sigma}_r$, RSD , S_r are synonymous.

REFERENCES FOR APPENDIX B

No.	Author	Title, etc
B1	W.R. Thompson	On a criterion for the rejection of observations and the distribution of the ratio of the deviation to the standard deviation. *An. Math. Stats.*, **6**, pp 214-219 (1935)
B2	F.E. Grubbs	Procedures for detecting outlying observations in samples. *Technometrics*, **11**, 1, February 1969
B3	P.G. Moore D.E. Edwards	Standard statistical calculations. Pitman and Sons (1965)
B4	G.L. Tietjens R.H. Moore R.J. Beckman	Testing for a single outlier in simple linear regression. *Technometrics, 15*, 4, November 1973
B5	R.B. Abernethy B.G. Ringhiser	Outlier rejection from regression models. Correspondence to ASME Com. PTC 19.1, 27 May 1981
B6	R.E. Lund	Tables for an approximate test for outliers in linear models. *Technometrics*, **17**,.4, November 1975

Table 1

EXAMPLE OF ELEMENTAL ERROR TABLE

(ONE BASIC MEASUREMENT)

ELEMENTAL ERRORS OF BASIC MEASUREMENT (LOAD)

LOAD = Test Frame Load		
Sensor = Bofors trunnion thrust load cell S/N 50745		
ERROR CATEGORIES	ERROR SOURCES	BIAS LIMITS [N]
1. CALIBRATION HIERARCHY	National force standards @ NPL	2
	NPL ↓	13
	HBM transfer std ↓	19
	Bofors pure curves	
2. DATA ACQUISITION	Small instrumentation errors included in Cat. 1 and 4	---
3. DATA REDUCTION	Non-linearity of load calibration	20
4. REAL EFFECTS	AV(z) bias ---- nil. with "scheme 2" OLZ reset	---
	Variation of reset OLZ,z′	25
	Hysterisis with closing throttle for target point	10
	RSS TOTALS	41

Table 2

EXAMPLE OF ERROR PROPAGATION TABLES
(TWO FLIGHT CONDITIONS)

ERROR PROPAGATION AT CONDITION 1, TARGET POINT (82.7 P2/1.0 RR/253 T2/8675 NHRD

	Basic measurements (input parameters) x_i (units)			Table No	Bias limits $\pm B_i$ (units)	(%)	Influence coefs θ_{ki} (%/%) SFCRD	FNRD	WAIRD	NFRD	Bias limits of results $\pm B_{ki} = B_i \cdot \theta_{ki}$ (%) SFCRD	FNRD	WAIRD	NFRD
1	PSA	74.22	kPa	11	0.044	0.06	-0.12	0.12	0.50	0	0.01	0.01	0.03	0
2	PA	33.93	kPa	12	0.035	0.25	-0.12	0.12	0.49	0	0.03	0.03	0.12	0
3	TA	253.1	K	17	1.0	0.40	0.12	-0.12	-0.49	0	0.04	0.04	0.20	0
4	T2	253.9	K	18	1.0	0.40	-0.62	0.12	0.50	-0.50	0.25	0.05	0.20	0.20
5	PS1	79.79	kPa	14	0.06	0.07	-1.89	1.93	0	0	0.13	0.14	0	0
6	PC	80.93	kPa	15	0.11	0.14	1.52	-1.50	0	0	0.21	0.21	0	0
7	LOAD	30620	N	8	41	0.13	-0.91	0.91	0	0	0.12	0.12	0	0
8A	QF	877.8	ml/s	4	1.1	0.12	1.00	0	0	1.0	0.12	0	0	0
8B	D15	0.8023	kg/l	5	0.0010	0.12	"	0	0	1.0	0.12	0	0	0.12
8C	NCV	43187	J/g	7	76	0.18	"	0	0	1.0	0.18	0	0	0.12
9	TP	290.6	K	6	0.9	0.35	-0.25	0	0	-0.25	0.09	0	0	0.09
10	CDA	0.985	—	13	0.0071	0.72	-0.25	0.25	1.00	0	0.18	0.18	0.72	0
11	BARO	98.63	kPa	9	0.13	0.13	-0.02	-1.17	-0.52	-1.19	0.00	0.15	0.07	0.15
12	CCP	-20	kPa.g	10	0.008	0.04	-0.004	-0.23	-0.10	-0.24	0.00	0.01	0.00	0.01
	P2	82.36	kPa	16	0.112	0.14	-1.24	0.24	-1.00	-1.00	0.17	0.03	0.14	0.14
							RSS TOTALS $\pm B_k$			% →	0.52	0.37	0.80	0.39

ERROR PROPAGATION AT CONDITION 2, TARGET POINT (82.7 P2/1.0 RR/268 T2/8775 NHRD)

1	PSA	75.16	kPa	11	0.044	0.06	-0.12	0.12	0.50	0	0.01	0.01	0.03	0
2	PA	12.93	kPa	12	0.033	0.26	-0.12	0.12	0.49	0	0.03	0.03	0.13	0
3	TA	268.0	K	17	1.0	0.37	0.12	-0.12	-0.49	0	0.04	0.04	0.18	0
4	T2	268.2	K	18	1.0	0.37	-0.62	0.12	0.50	-0.50	0.23	0.04	0.18	0.18
5	PS1	80.07	kPa	14	0.06	0.07	-2.12	2.16	0	0	0.15	0.15	0	0
6	PC	81.06	kPa	15	0.11	0.14	1.69	-1.67	0	0	0.23	0.23	0	0
7	LOAD	27479	N	8	41	0.15	-0.90	0.91	0	0	0.14	0.14	0	0
8A	QF	806.8	ml/s	4	0.97	0.12	1.00	0	0	1.0	0.12	0	0	0
8B	D15	0.8023	kg/l	5	0.0010	0.12	"	0	0	1.0	0.12	0	0	0.12
8C	NCV	43187	J/g	7	76	0.18	"	0	0	1.0	0.18	0	0	0.12
9	TP	287.5	K	6	0.9	0.35	-0.25	0	0	-0.25	0.09	0	0	0.09
10	CDA	0.985	—	13	0.0071	0.72	-0.25	0.25	1.00	0	0.18	0.18	0.72	0
11	BARO	98.63	kPa	9	0.13	0.13	-0.01	-1.17	-0.53	-1.18	0.00	0.15	0.07	0.15
12	CCP	-20	kPa.g	10	0.008	0.04	-0.002	-0.23	-0.11	-0.24	0.00	0.01	0.00	0.01
	P2	82.38	kPa	16	0.112	0.14	-1.24	0.24	-1.00	-1.00	0.17	0.03	0.14	0.14
							RSS TOTALS $\pm B_k$			% →	0.53	0.39	0.79	0.38

GLOSSARY

Note: This Glossary is intended for quick reference with enough explanation to act as a convenient reminder of the main features of the methodology.

Accuracy	A conversational type of word used to convey a qualitative impression of the goodness of measurements and results. However, the word is so popular and has so many different meanings, that it is now too confusing to use in a technical sense.
Actual Results Analysis	The statistical analysis of the random scatter about the test result, usually by analysis of a performance curve to calculate its Residual Standard Deviation (RSD) and Random Error Limit of Curve Fit (RELCF). This involves only the precision errors which appear with the test result. Bias errors, being constant, are not revealed.
Average	Synonymous with 'mean value'. The random uncertainty of an average of 'n' points is $1/\sqrt{n}$ times the uncertainty of a single point, *ie* it is thereby improved.
Basic Measurement, xi	(Sometimes called an 'Input Parameter'.) One of a number of variables (such as airmeter static pressure, PSA, or fuel flow rate, QF, etc) which appear in the mathematical expressions through which a test result is calculated from the measurements. A separate Elemental Error Table is published for each Basic Measurement - see Table 1 for an example.
Bias (or Systematic) Error, β	A constant error which lies somewhere within the limits $\pm B$. The value of the error itself is unknown but the limits can be estimated by 'Prediction Synthesis'.
Bias Limits, $\pm B$	The maximum plus/minus values within which a bias error is expected to lie.
Curve Fitting	Synonymous with 'regression', a process by which a best curve, of some pre-defined mathematical form, is fitted to the test data by the method of least squares. The statistical properties of a curve fit are similar to those of a mean value.

GLOSSARY (continued)

Curve Shift Effect — Because engine performance results are displayed as graphical correlations against other results, it is necessary to account for the propagation of errors not only directly to the vertical ordinate, y, but also indirectly via the horizontal abscissa, z. Bias errors in z do not affect the y-value of a point but they do shift the position of the curve (see Fig 17).

Defined Measurement Process — This embraces every factor and every measurement that has an influence upon the 'Defined Test Result', not only in the test itself, but also stretching all the way back through the calibration hierarchy to the National Standards.

Defined Test Result — In ordinary usage, the 'Test Result' would be an engine performance parameter (such as Net Thrust, FN or Specific Fuel Consumption, SFC, etc) the determination of which is the object of the test. In this document, the 'Defined Test Result' has an extra special meaning for the purpose of uncertainty assessment. For this it is necessary to declare exactly what form the result is to be:

either a single point value, R
or a mean value, $\bar{R}$
or a curve fit value, $\hat{R}$.

Degrees of Freedom, υ — The number of independent points of data involved in a statistical calculation. For a mean value, $\upsilon = n - 1$. For a curve fit $\upsilon = n - p - 1$, where p is the power of the polynomial curve.

Elemental Errors — These are the most fundamental errors that can affect each Basic Measurement. They are accounted as various Error Sources in four Error Categories.

Error Categories, j — For each Basic Measurement, there are four categories into which it is convenient to group the various Error Sources for accounting purposes, thus

(1) Calibration hierarchy
(2) Data acquisition
(3) Data reduction
(4) Real effects.

GLOSSARY (continued)

Error Classification — It is first necessary to declare exactly what is meant by the 'Defined Test Result', for example an SFC curve. Then any error which produces random scatter in that result is classified as 'Precision', and any error which remains constant (and hence unseen) in that result is classified as 'Bias'.

Error Propagation — The Bias Limits of each Elemental Error within each Error Category, j, are first assessed and accounted in a separate table for each Basic Measurement, x_i. These are combined separately by root-sum-squares, thus

$$B_i = \sqrt{\sum_j B_{ij}^2}$$

and recorded at the bottom of the separate tables (*eg* see Table 1). These B_i values for each Basic Measurement, xi, are then propagated to the Test Result, R_k, by means of the appropriate Influence Coefficients, and again combined by root-sum-squares, thus

$$B_k = \sqrt{\sum_i \left[\theta_{ik} B_i\right]^2}$$

to give the total effect on each test result (*eg* see Table 2).

In principle, similar prediction synthesis could be applied to the Precision Indices but, in RAE practice at Pyestock, it is considered better to assess this by Actual Results Analysis (qv).

Error Sources — These are the most fundamental origins to which the Elemental Errors can be ascribed, within each Error Category.

Fossilisation — Random errors which occur before the start of the engine test run, for example during an instrument calibration, will leave a small fixed error in the calibration curve. This fixed error is then propagated to the engine test. Thus such effects of random errors are 'fossilised' into bias errors in the engine test result.

GLOSSARY (continued)

Gaussian Distribution

(The alternative name is 'Normal Distribution'.)
The bell-shaped probability distribution of random variables, which is described in all text books of statistics. The most likely value is the mean value, while 95% of values are found within $\pm 2\sigma$, where σ is the theoretical standard deviation. In practical terms, σ is replaced by the Precision Index, s, or by the Residual Standard Deviation, RSD.

Influence Coefficient, θ_{ik}

The rate of change of a Test Result, R_k, with respect to a Basic Measurement, x_i

$$ie \quad \theta_{ik} = \frac{\partial R_k}{\partial x_i}$$

$$or \quad \theta'_{ik} = \frac{x_i}{R_k} \frac{\partial R_k}{\partial x_i}$$

in relative, or percentage form.

Precision (or Random) Error, ε

A random variable, assumed to come from a Gaussian distribution, which lies somewhere within the limits $\pm ts$, or $\pm tRSD$. The presence of such errors is shown by the scatter of points about a mean value, or about a curve fit. The value of 's' or 'RSD' is found by Actual Results Analysis, and Student's "t" is extracted from statistical tables.

Precision Index, s

The experimental standard deviation of the Gaussian distribution from which the Precision Errors arise. In principle, the value of 's' could be estimated by Prediction Synthesis but it is RAE(P) practice to calculate it from statistical analysis of actual data.

Prediction Synthesis

This is important as the only way to assess the effects of bias errors. (Precision errors can also be included but these are best left to actual results analysis.) Every error source in the defined measurement process is examined and assessed for elemental bias limits, with a separate table for each basic measurement (or input parameter). These bias limits are then propagated to the test result by means of influence coefficients, and combined to predict the overall effect.

GLOSSARY (continued)

Probability — The likelihood of a value being realised. The word 'probability' is synonymous with 'relative frequency'.

Probability Distribution — In diagrammatic form, the horizontal axis shows all possible values of a variable (x , say) while the vertical axis shows the 'probability density' of x , denoted as dP/dx (*eg* Figs A1 and A2 of Appendix A). The probability of a value of x occurring between A and B is given by

$$\int_{B}^{A} \frac{dP}{dx}\, dx \ ,$$

ie the area under the curve between A and B . The total area under the curve

$$\int_{-\infty}^{\infty} \frac{dP}{dx}\, dx$$

is equal to 1, *ie* total probability.

RELCF — The random error limits of curve fit which describe a band enclosed by the limits ±RELCF on either side of a curve fit. If there are no bias errors, the true curve will be expected to lie somewhere within this band, at 95% confidence. The statistical properties of RELCF are similar to those of a mean value, *ie* the band ±RELCF is narrower than the scatter of individual points.

RELIP — The random error limits of individual points scattered about a curve fit. These limits are approximated by ±2RSD.

Residual Standard Deviation, RSD — The standard deviation of the residual scatter of points about a curve fit. It is calculated by statistical methods in Actual Results Analysis.

Student's "t" — A factor, extracted from statistical tables against the appropriate degrees of freedom, to give the confidence limits ±ts, or ±tRSD, within which a precision error is expected to lie. The confidence level is usually taken to be 95%.

GLOSSARY (concluded)

Uncertainty — This word is used instead of 'Accuracy' in a technical treatment, such as the present document. Its numerical value is the half-range of an interval within which the error is expected to lie. Uncertainty has two components of error, bias and precision, which may be combined in alternative ways. Thus, for the Uncertainty of a single point,

$$U_{add} = B + 2s$$

or

$$U_{rss} = \sqrt{B^2 + (2s)^2} \ .$$

(Note: The Precision Index, s , could be replaced by the RSD.)

NOTATION

(see Glossary for detailed explanations)

Roman symbols	Description
B	bias limit
BLCF	bias limit of curve fit
CF	curve fit
CL_{95}	95% confidence limit
DMP	defined measurement process
FNRD	net thrust, referred to desired conditions
G	a number from a Gaussian distribution
n	number of points
NHRD	high pressure shaft speed, referred to desired conditions
P	probability (*ie* relative frequency)
P7Q2	engine pressure ratio = P7/P2
R	either (a) a test result or (b) a number from a random rectangular distribution
RELCF	random error limit of curve fit
RELIP	random error limit of individual points
RSD	residual standard deviation
s	precision index (*ie* experimental sample standard deviation)
SFCRD	specific fuel consumption, referred to desired conditions
TV	true value
t	Student's "t" factor, usually for 95% confidence
U_{add}	additive uncertainty = (B + ts)
U_{rss}	root-sum-square uncertainty = $\sqrt{B^2 + (ts)^2}$
UETP	uniform engine test programme
U_r	random uncertainty = ts
WA1RD	engine inlet air flow rate, referred to desired conditions
x	a basic measurement
y	a main test result, plotted against graph ordinate
z	a secondary test result, plotted against graph abscissa (*ie* a correlating result)

NOTATION (concluded)

Greek symbols	Description
β	a bias (systematic) error
ε	a precision (random) error
θ_{ki}	influence coefficient of a result R_k to a basic measurement, xi $= \frac{\partial R_k}{\partial x_i}$
σ	theoretical population standard deviation

Superscripts

-	mean value
^	curve fit value

Suffixes

i	general basic measurement
j	general source of elemental error
k	general test result

REFERENCES

No.	Author	Title, etc
1	R.B. Abernethy J.W. Thompson	Handbook: Uncertainty in gas turbine measurements. AEDC-TR-73-5, April 1973
2	J.C. Ascough	Prediction synthesis of error limits for J57 UETP tests in Cell 3, Pyestock. RAE Technical Memorandum P 1069 (1985)
3	J.C. Ascough	A preliminary post test analysis of bias errors between UETP test facilities. RAE Technical Memorandum P 1068 (1985)
4	J.C. Ascough	Uncertainty assessment in the Uniform Engine Test Programme. RAE Technical Memorandum P 1165 (1989)
5	I.A. Harris	The expression of uncertainty in electrical measurements. British Calibration Service Document 3003, April 1977
6	MIDAP Study Group	Guide to in-flight thrust measurement of turbo-jets and fan engines. Chap. 4: Error assessment and control. AGARD-AG-237, January 1979
7	M.J. Moroney	Facts from figures. Penguin Books Ltd, 3rd Ed. (1965)
8	O.L. Davies P.L. Goldsmith	Statistical methods in research and production. Oliver and Boyd, 4th Ed. (1972)
9	C.A. Bennett N.L. Franklin	Statistical analysis in chemistry and the chemical industry. John Wiley and Sons Ltd, 5th printing (1967)
10	J.C. Ascough	A Test Code for contract performance measurements made in Cell 3 of the NGTE engine test facility. NGTE NT 1088, April 1977
11	J.C. Ascough	Allowance for the effect of small deviations of inlet air temperature on engine thrust at defined ratings. NGTE M78021, May 1978

REPORT DOCUMENTATION PAGE

Overall security classification of this page

UNLIMITED

As far as possible this page should contain only unclassified information. If it is necessary to enter classified information, the box above must be marked to indicate the classification, e.g. Restricted, Confidential or Secret.

1. DRIC Reference (to be added by DRIC)	2. Originator's Reference	3. Agency Reference	4. Report Security Classification/Marking
	RAE TM P 1179		UNLIMITED

5. DRIC Code for Originator	6. Originator (Corporate Author) Name and Location
7674300E	Royal Aerospace Establishment, Pyestock, Hants, UK
5a. Sponsoring Agency's Code	6a. Sponsoring Agency (Contract Authority) Name and Location

7. Title Handbook of uncertainty methodology for engine testing at Pyestock

7a. (For Translations) Title in Foreign Language

7b. (For Conference Papers) Title, Place and Date of Conference

8. Author 1. Surname, Initials	9a. Author 2	9b. Authors 3, 4	10. Date	Pages	Refs.
Ascough, J.C.			November 1989	65	11

11. Contract Number	12. Period	13. Project	14. Other Reference Nos.

15. Distribution statement

(a) Controlled by –

(b) Special limitations (if any) –

If it is intended that a copy of this document shall be released overseas refer to RAE Leaflet No.3 to Supplement 6 of MOD Manual 4.

16. Descriptors (Keywords) (Descriptors marked * are selected from TEST)

Uncertainty, Engine testing, Statistics, Curve fitting, Error prediction, Great Britain. (SDW)

17. Abstract

Measurements of aircraft gas turbine engine performance in the altitude test facility at RAE(P) are subject to a small amount of uncertainty resulting from a combination of precision (or random) errors and bias (or systematic) errors. The limits of the precision errors can be readily calculated by statistical analysis of the results measured during the engine tests. Bias limits are not directly observable in the test results, but can be predicted by a comprehensive assessment of all possible sources of error, which are propagated to the test results.

Many people find these methods difficult to comprehend and apply and this Memorandum has been written for their benefit. It is a guide not only for engine test staff at Pyestock but also for their customers who need to be assured of the rigorous attention given to identifying and reducing measurement uncertainty.

Keywords

F5910/1

RAE Form A143 (revised October 1980)